Psychological Mechanisms of Loss Aversion: A Drift-Diffusion Decomposition

Wenjia Joyce Zhao

University of Pennsylvania

Lukasz Walasek

University of Warwick

Sudeep Bhatia

University of Pennsylvania

July 1, 2020

Send correspondence to Wenjia Joyce Zhao, Department of Psychology, University of Pennsylvania, Philadelphia, PA. Email: zhaowenj@sas.upenn.edu

Abstract

Decision makers often reject mixed gambles offering equal probabilities of a larger gain and a smaller loss. This important phenomenon, referred to as loss aversion, is typically explained by prospect theory, which proposes that decision makers give losses higher utility weights than gains. In this paper we consider alternative psychological mechanisms capable of explaining loss aversion, such as a fixed utility bias favoring rejection, as well as a bias favoring rejection prior to gamble valuation. We use a drift diffusion model of decision making to conceptually distinguish, formally define, and empirically measure these mechanisms. In two preregistered experiments, we show that the pre-valuation bias provides a very large contribution to model fits, predicts key response time patterns, reflects prior expectations regarding gamble desirability, and can be manipulated independently of the valuation process. Our results indicate that loss aversion is the result of multiple different psychological mechanisms, and that the pre-valuation bias is a fundamental determinant of this well-known behavioral tendency. These results have important implications for how we model behavior in risky choice tasks, and how we interpret its relationship with various psychological, clinical, and neurobiological variables.

Keywords: loss aversion, risky choice, decision making, drift diffusion model, computational modeling

1. Introduction

Consider a gamble that offers a gain of $11 if a coin toss lands heads, and a loss of $10 if it lands tails. Would you accept or reject this gamble? Most people choose to reject similar positive-expected-value mixed gambles (gambles that have the potential for both gains and losses; Kahneman & Tversky, 1979; P. A. Samuelson, 1960; Tversky & Kahneman, 1992). According to expected utility theory, this simply reflects concavity of utility functions, which leads to risk aversion. Yet rejection rates observed for mixed gambles involving such small monetary payoffs cannot be easily explained by conventional applications of expected utility theory. As Rabin (2000) points out, for such models to predict the rejection of a 50-50 gamble between a gain of $11 and a loss of $10, the assumed degree of risk aversion would have to be so high that an individual would reject any 50-50 gamble involving a loss of $100, regardless of the magnitude of the corresponding gain. This unreasonable prediction presents compelling evidence against risk aversion being the only cause of mixed gamble rejection, and suggests that people are *loss averse*, that is they display an additional (psychological) aversion to gambles that offer the possibility of a loss.

1.1. Prospect theory explanations for loss aversion

Loss aversion has commonly been understood through the lens of prospect theory (Kahneman & Tversky, 1979), which states that losses are valued differently than gains[1] (also see Kahneman, 2003; Kőszegi & Rabin, 2007; Rabin & Thaler, 2001; Tversky & Kahneman, 1992). Specifically, prospect theory proposes that decision makers place a greater utility weight

[1] In this paper we distinguish between *loss aversion* (which refers to the general psychological tendency to avoid gambles that offer the possibility of a loss) and prospect theory's account of loss aversion (which takes the form of biased utility weights, specified using the λ parameter discussed below).

on losses and a lower utility weight on gains[2]. Thus, in the decision to accept or reject a gamble i, offering a 50% chance of gaining G_i and a 50% chance of losing L_i ($G_i, L_i > 0$), the utility for accepting the gamble, according to prospect theory, is given by $U_i = G_i - \lambda \cdot L_i$ (as the probabilities of the gains and losses are identical, we assume that their effects can be ignored without influencing model predictions, but also see Tversky & Kahneman, 1992 and Pachur & Kellen, 2013 on applying different probability weighting functions in the gain and loss domains). Here λ is the prospect theory loss aversion parameter, where $\lambda > 1$ indicates that losses have a larger utility weight than gains. Assuming that the utility for rejecting the gamble is 0, the decision maker will accept the gamble when $U_i > 0$, and reject the gamble when $U_i < 0$. In the gamble presented at the start of this paper, a large enough value of λ implies that individuals experience more negative utility from the loss of \$10 than positive utility from the gain of \$11. Thus, the gamble, despite having a positive expected value, appears unattractive and is rejected.

As prospect theory has become the predominant theory for describing choice under risk and uncertainty, its utility weighting explanation (i.e. the assumption that $\lambda > 1$) has also become widely accepted as *the psychological mechanism* responsible for loss aversion. Indeed, an individual's degree of loss aversion[3], measured by their gamble rejection rate, is often seen to be synonymous with the value of λ. This assumption is, in turn, used to relate prospect theory to various behavioral, cognitive, clinical, demographic, and neurobiological variables. Among many examples of this approach, researchers have argued that biased utility weights (in the form

[2] In this paper we use *utility weights* to refer to the weighting coefficients for gains and losses in various forms of value functions. This is not to be confused with decision weights in the probability weighting function of prospect theory. In this paper, we only consider mixed gambles with equal probabilities of gains and losses.

[3] If prospect theory utility, $U_i = G_i + \lambda \cdot L_i$, was the only determinant of gamble acceptance or rejection, λ would be highly correlated with the rejection rate and would be *the* explanation of loss aversion. However, as we discuss below, there are many alternate explanations for gamble rejection.

of $\lambda > 1$) play an important role in irrational financial decision making, problem gambling, suicidal decision making, and incorrect affective forecasting (Benartzi & Thaler, 1999; Hadlaczky et al., 2018; Kermer, Driver-Linn, Wilson, & Gilbert, 2006; Lorains et al., 2014; Takeuchi et al., 2015); in explaining differences in risky decision making between decision contexts (Polman, 2012; Schulreich, Gerhardt, & Heekeren, 2016; Sokol-Hessner, Camerer, & Phelps, 2013; Sokol-Hessner, Raio, Gottesman, Lackovic, & Phelps, 2016; Vermeer, Boksem, & Sanfey, 2014) and between individuals with varying psychological traits, demographic profiles, and life experiences (Barkley-Levenson, Van Leijenhorst, & Galván, 2013; Barkley-Levenson & Galvan, 2014; Bibby & Ferguson, 2011; Pighin, Bonini, Savadori, Hadjichristidis, & Schena, 2014; Sokol-Hessner, Hartley, Hamilton, & Phelps, 2015a); and in determining physiological and neural responses to risky prospects (Canessa et al., 2017; 2013; De Martino, Camerer, & Adolphs, 2010; Engelmann, Meyer, Fehr, & Ruff, 2015; Gelskov, Henningsson, Madsen, Siebner, & Ramsøy, 2015; Lazzaro, Rutledge, Burghart, & Glimcher, 2016; Markett, Heeren, Montag, Weber, & Reuter, 2016; Sokol-Hessner et al., 2013; Sokol-Hessner, Lackovic, Tobe, Camerer, Leventhal, et al., 2015b; Takahashi et al., 2013; Tom, Fox, Trepel, & Poldrack, 2007). A single prolific and influential example of this approach is presented in Tom et al. (2007): In this paper, neural activity was correlated with gamble rejection rates, which the authors interpreted as identifying brain regions that encode values of prospect theory's λ parameter.

1.2. Alternate explanations for loss aversion

Recent work has challenged the utility weighting explanation proposed by prospect theory[4] (see review by Gal & Rucker, 2018; but also see Simonson & Kivetz, 2018). Rather than

[4] Note that these debates do not only apply to research on gambles, but also to other topics such as the endowment effect and hedonic impact ratings. As our focus is on risky choice, we limit our discussion to the most relevant evidence in this area.

a universal property of people's preferences, utility weights for losses and gains may be context-dependent and can be absent or even reversed in certain decision environments. For example, decision makers no longer place higher weights on losses than gains when the accept-reject paradigm is framed as an equivalent binary choice between a mixed gamble and a sure outcome of $0 (Ert & Erev, 2013). The multiplicative effect of losses on utility can even be of a smaller magnitude than that of gains when individuals experience a wider range of losses compared to gains in repeated trials (Walasek & Stewart, 2015; Walasek & Stewart, 2019).

It is also the case that current approaches to modeling loss aversion using prospect theory utility weights implicitly allow for other mechanisms to influence choice. Recall that the utility for accepting a gamble *i*, according to prospect theory, is given by $U_i = G_i - \lambda \cdot L_i$. Stochasticity in choice can be modeled with a logistic response function. With such specification, the magnitude of λ can be estimated using a logistic regression: $A_i \sim \beta_G \cdot G_i - \beta_L \cdot L_i$. Here A_i is the participant's binary response to the gamble (1 if Accept, 0 if Reject), and β_G and β_L are regression coefficients that yield the utility weighting effect as specified by prospect theory with $\lambda = \beta_L/\beta_G$. In practice, researchers often include an additive intercept, α, in the utility function, i.e., $U_i = \alpha + G_i - \lambda \cdot L_i$. Correspondently, the logistic regression for predicting choice becomes $A_i \sim \alpha + \beta_G \cdot G_i - \beta_L \cdot L_i$ (for a discussion of this point see Walasek & Stewart, 2019). This additive intercept generates a fixed utility bias favoring rejection ($\alpha < 0$) or acceptance ($\alpha > 0$). Unlike utility weights, the magnitude of this fixed utility bias, and the effect it generates, is not dependent on the specific gain and loss amounts offered by the gamble. When α is negative, decision makers can reject positive-EV mixed gambles even without higher utility weighting for losses.

The fixed utility bias (α) and the prospect theory utility weighting bias ($\lambda = \frac{\beta_L}{\beta_G} > 1$) make up the decision maker's utility function. Yet, individuals can also exhibit loss aversion *prior to* gamble valuation; that is, individuals could be predisposed to rejection even before they have inspected and learnt about the monetary amounts that could be gained or lost. Psychologically, this pre-valuation bias can be seen as a type of status-quo bias or psychological inertia (Gal, 2006; Samuelson & Zeckhauser, 1988; Gal & Rucker, 2018), a tendency according to which individuals avoid actions that lead to potential losses relative to the current state of affairs (though, as we discuss below, there are other compelling statistical and neurobiological interpretations of this mechanism). Although such a tendency may be overridden after the gamble is valued, we would nonetheless expect the pre-valuation bias to influence people's decisions and, in many settings, lead to a higher probability of rejection than acceptance.

Considerable research has modeled risky choice with the assumption of flexible utility weights (and often, implicitly, additive intercepts in utilities – e.g. Pachur & Scheibehenne, 2017; Schulreich, Gerhardt, & Heekeren, 2016; Stewart, Reimers & Harris, 2015; Walasek & Stewart, 2015). Yet there has been almost no work that has also allowed for the effect of a pre-valuation bias on choices among mixed gambles. This is largely because mechanisms such as the pre-valuation bias cannot be accommodated within the types of economic models used to predict risky choice. Typically, these economic models assume that choices depend entirely on *utility*, which itself is a product of the gains and losses offered by the gamble in consideration. Thus, there is no place for a psychological mechanism for loss aversion that influences choice *prior to* the formation of utility.

Due to this technical constraint, previous discussions of pre-valuation biases, and related mechanisms, appear almost only in verbal models, and this inevitable lack of formal definition

and quantitative measurement severely limits the scope of analyses performed in theoretical and empirical research. Thus, for example, prior research documenting correlations between gamble rejection rates and various psychological, clinical, and neurobiological variables, may have misattributed these correlations to differences in λ across decision environments and individuals, when they may be better understood in terms of differences in the pre-valuation bias. Similarly, the pre-valuation bias could explain why people display less loss aversion when the accept-reject decision is framed as an equivalent binary choice between a mixed gamble and a sure outcome of $0 (Ert & Erev, 2008, 2013; see also Erev et al., 2008). As the pre-valuation bias predisposes the decision maker to rejection over acceptance, its effect is likely to be larger in the accept-reject paradigm than the binary choice paradigm. More generally, allowing for the pre-valuation bias in a formal model of risky choice can shed light on a fuller set of psychological mechanisms responsible for loss aversion, and by doing so, allow for improved predictions of behavior, and a more sophisticated understanding of individual differences, contextual influences, and the effects of various psychological, clinical, and neurobiological variables on risky choice.

2. Model

Some researchers have turned to models originally proposed in mathematical and cognitive psychology to develop a comprehensive modeling framework for decomposing the numerous psychological mechanisms responsible for decision under risk (e.g., Bhatia, 2014; Busemeyer & Townsnd, 1993 Clay, Clithero, Harris & Reed, 2017; Diederich & Trueblood, 2018; Pleskac, Wallsten, Wang & Lejuez, 2008; Rieskamp, 2006; Trueblood, Heathcote, Evans & Holmes, 2019). In this paper, we study decision processes in mixed gamble decisions using one such model: the drift diffusion model (DDM) – a sequential sampling model which assumes that individuals gradually accumulate evidence over the time course of the decision, with a

decision being made when evidence reaches a threshold value (Ratcliff, 1978). The DDM, and related models, have proved successful in accounting for choice and response time data observed in a wide range of non-risky perceptual and preferential choice tasks (Dai & Busemeyer, 2014; Krajbich, Armel, & Rangel, 2010; Ratcliff, 1978; Ratcliff & McKoon, 2008; Tsetsos, Chater, & Usher, 2012; Tsetsos et al., 2016; White & Poldrack, 2014; Zhao et al., 2019), and provide the leading quantitative framework for modeling the numerous psychological influences at play in two-option forced choice, such as the accept-reject decisions studied in this paper.

As is illustrated in Figure 1A, we assume that for mixed gamble choice, decision makers' starting point of preference accumulation is γ. Once the accumulation of evidence begins, they integrate evidence in favor of accepting vs. rejecting the gamble over time, with a drift rate (v) that relates the utility of the gamble to the accumulation process. Choices are made when the accumulated evidence reaches a positive threshold $+\theta$ (corresponding to acceptance) or a negative threshold $-\theta$ (corresponding to rejection). The response time (RT) in a trial is assumed to be the time taken for the accumulating evidence to reach a decision threshold added to a fixed non-decisional time τ (which captures the time taken to perceive the stimuli, execute motor responses after the decision has been made, and so on). We do not assume any between-trial variability in the parameters.

In perceptual choice, the speed of evidence accumulation is dependent on the signal strength, e.g., consistency of random dot movement or contrast between line segments. In preferential choice, the evidence being accumulated depends on features of the choice alternatives, and subsequently their relative utilities. Following this logic, the drift rate for a trial involving gamble i is given by $v_i = U_i(\text{Accept}) - U_i(\text{Reject})$. To keep model specifications consistent with the logistic model introduced in Section 1, we write the drift rate as $v_i = \alpha +$

$\beta_G \cdot G_i - \beta_L \cdot L_i$. Note that in the absence of a pre-valuation bias, which we will soon introduce, the DDM has the same functional form for predicting choice probabilities as the static logistic model outlined above. In other words, the static logistic model used in prior work can be seen as a special case of the DDM, as it imposes additional parameter constraints (no pre-valuation bias) and does not predict RTs. Therefore, the ratio between the loss vs. gain coefficient in the drift rate, β_L/β_G, is the DDM-equivalent measure of the prospect theory utility weighting bias, λ, whereas α is a DDM-based measure of the fixed utility bias (i.e., the intercept in a logistic regression). In cases when $\lambda = \frac{\beta_L}{\beta_G} > 1$ and/or $\alpha < 0$ one may observe a higher propensity to reject mixed gambles.

Beyond incorporating previously proposed mechanisms into a dynamic modeling framework and thus potentially providing more accurate measures of their effects (Clithero, 2018), our approach is able to define and measure a new mechanism. This new mechanism, the pre-valuation bias, takes the form of a starting point (γ) in the drift diffusion model. A positive γ, that is closer to $+\theta$, predisposes the decision maker towards accepting the gamble. A negative γ , that is closer to $-\theta$, predisposes the decision maker towards rejecting the gamble. When $\gamma = 0$, the preference accumulation process starts from a neutral state[5].

Apart from offering conceptual *definitions* for distinct biases, the DDM also *predicts* that the pre-valuation bias will generate a unique *behavioral marker* regarding the relationship between choice and response time, that cannot be explained using a fixed utility bias or the

[5] Note that the idea of a stimulus-independent bias is not new. Indeed, many researchers have included an additive intercept in a logistic regression analyses to capture a general tendency of choice making, independently of the monetary amounts on offer (e.g., Pachur & Scheibehenne, 2017; Schulreich, Gerhardt, & Heekeren, 2016; Stewart, Reimers & Harris, 2015; Walasek & Stewart, 2015). However, this term is conceptually similar to the intercept in the drift rate (α) which influences decision making in parallel to the stimuli-dependent valuation of the gamble. Indeed, as we discuss above, the DDM reduces to the logistic regression model used in this prior work when the pre-valuation bias is zero.

utility weighting bias. Without a pre-valuation bias, the DDM predicts that RT distributions associated with acceptance vs. rejection decisions for a gamble should be identical. Imposing a pre-valuation bias that sets a starting point closer to the rejection threshold would make the choice of rejection quicker, compared to the choice of acceptance. Intuitively, this is because the pre-valuation bias is implemented before the valuation process starts and its effect on choice and RTs diminishes as the decision maker deliberates about the money that could be gained or lost. In contrast, drift rate effects persist throughout the decision process (see prior discussions on the separation between the starting point parameter and the drift rate parameters in a drift diffusion model: e.g., Leite & Ratcliff, 2011; Voss, Rothermund, & Voss, 2004; Voss, Voss, & Klauer, 2010; White & Poldrack, 2014).

In addition to the above prediction, the various parameters of our DDM implementation of loss aversion can also be quantitatively estimated and differentiated from one another with a combination of choice and response time data. In prior work, psychologists and neuroscientists have used these estimates to compare pre-valuation biases against alternative decision mechanisms in a variety of perceptual and preferential choice tasks (e.g., White & Poldrack, 2014; Zhao et al., 2019).

The pre-valuation bias also has compelling statistical and neurocognitive interpretations. Mathematically, the DDM implements a sequential probability ratio test, which achieves a pre-selected level of accuracy in the shortest possible time. With this interpretation, a starting point bias in the accumulation process is related to unequal prior expectations regarding how rewarding different responses are. If decision makers were behaving adaptively, the starting point should be closer to the choice boundary of the response with a higher prior probability of being more rewarding. In line with this normative prediction, many perceptual experiments

demonstrate that the effect of manipulating prior probabilities or reward values for responses can be described by a shift in the participant's starting point biases (e.g., Ratcliff & Smith, 2004; Leite & Ratcliff, 2011). For example, in a word recognition task where participants classify words into two categories ("old" or "new"), researchers manipulate prior probabilities by making "old" the correct answer in more than 50% questions. Alternatively, researchers manipulate reward values by allocating higher reward amounts for the correct response to the "old" words, compared to those to the "new" words. In both cases, participants experience "old" as the more rewarding response, and develop a starting point bias that favors choosing "old". The DDM has also been suggested as a simplified model of neural information processing. With this interpretation, a starting point bias can be seen as a bias in baseline firing rates of neural units. In other words, a starting point bias causes pre-evaluation response tendencies inclining the decision maker towards one type of response (Bogacz, 2007; Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Gold & Shadlen, 2007; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012).

In the present paper we apply the DDM to the context of risky choice. Building on the findings discussed above, we use the DDM to decompose the psychological mechanisms that contribute to high rejection rates of mixed gambles (see Table 1 for a summary). Our objective is to determine the extent to which participants display pre-valuation biases towards rejecting (or accepting) mixed gambles. We also hope to shed light on the importance of this mechanism, compared to the dominant explanation in the field – prospect theory utility weighting captured by the parameter λ. Extending past work on risky choice and loss aversion, our approach also allows us to test whether biases towards accepting and rejecting gambles adapt dynamically to previous experience, that is whether it is possible to alter the pre-valuation bias by altering the gambles

the decision maker is exposed to. Thus, participants expecting to reject more than 50% of the trials prior to gamble evaluation may begin their decision with a starting point favoring rejection, whereas participants expecting to accept more than 50% of the trials prior to gamble evaluation may begin their decision with a starting point favoring acceptance. If the pre-valuation bias can in fact be shifted by manipulating payoff distributions, it can be interpreted as a consequence of prior expectations that rejection is more likely to be the more rewarding response. Such expectations may have been formed based on prior experiences in the lab or in the real-world, and would reflect an adaptive approach to making accept-reject decisions involving mixed gambles.

3. Experiment 1

In this experiment we presented participants with gambles involving equal probabilities of monetary gains and losses. Their task was to indicate whether they would accept or reject the gambles. We fit DDMs to choice and RT data so as to decompose participants' responses into various underlying psychological mechanisms, and infer the existence and relative importance of the prospect theory utility weighting bias for losses over gains, the fixed utility bias, and the pre-valuation bias. We preregistered our experimental methods, model specifications and main hypotheses at https://osf.io/varx6.

3.1. Methods

3.1.1. Procedures

49 participants (age: mean $= 22.55$, SD $= 6.10$; 67.3% female) were recruited from a university experimental research subject pool and performed the experiment on computers in a laboratory.

They were instructed to accept or reject a sequence of 200 gambles, presented in four blocks of 50 gambles each. Each gamble had two possible outcomes: A gain of some number of tokens occurring with a 50% chance and a loss of some number of tokens occurring with a 50% chance. The outcomes were displayed side by side in counterbalanced positions, with positive/negative values indicating gains and losses (see Figure 1B). Participants pressed up or down arrows on a keyboard to indicate acceptance or rejection, with the specific key-response associations alternating across blocks to control for response biases favoring one of the keys. Choices and response times were recorded.

Each token was worth US$0.10, and participants began the experiment with an endowment of 100 tokens (US$10). Participants were informed that their choices in the experiment would determine their bonus payment, which they would receive on top of a fixed show-up fee of US$8. This was accomplished by selecting one of the gambles at random at the end of the experiment. If the participant rejected the gamble, the bonus payment would be 100 tokens (US$10). If the participant accepted the gamble, then they would flip a coin in front of the experimenter to play out the gamble. Their received token amount would be their initial endowment (100 tokens = US$10) plus or minus the gain or loss associated with the coin flip.

3.1.2. Stimuli

The possible gain and loss values were taken from the set of {10, 20, 30, 40, 50, 60, 70, 80, 90, 100} tokens, or equivalently US$ {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}. With this stimulus set we were able to generate a total of 100 unique gambles. We counterbalanced the positions of the gain and loss outcomes for gambles, resulting in 200 total trials (see Figure 1C for an illustration of the payoff distributions).

3.2. Descriptive results

Consistent with prior work, we found that decision makers were mostly unwilling to accept the mixed gambles offered in the experiment. Overall, the average rejection probability across participants was 71.5% (SD = 0.19), with 79.6% of participants being more likely to reject than accept the gambles. These probabilities are larger than 50%, which is the rate we would expect if choices were made by chance or if individuals did not display any loss aversion ($t(48) = 8.03, p < 0.001$ when compared to 50%). Figure 2A (left panel) presents average acceptance rates as a function of the ratio of gains to losses, and indicates that, on average, participants accepted the gambles only when the size of the gain exceeded 1.75 times the size of the loss. Do note that the overall choice probability pattern is compatible with the prediction of a pre-valuation bias towards rejection ($\gamma < 0$), a fixed utility bias favoring rejection ($\alpha < 0$), and/or a prospect theory weighting bias for losses ($\lambda > 1$).

We also found that rejections were quicker than acceptances. Overall, the average rejection decision took 1.30 seconds, whereas the average acceptance decision took 1.72 seconds ($\Delta\text{M} = 0.41$ seconds, $t(46) = 4.05, p < 0.001$). Additionally, 74.5% of participants took less time to reject than to accept. Figure 2B (left panel) plots the overall distribution of mean response times for acceptance and rejection decisions across participants, and shows that these distributions are different from each other (Wilcoxon signed rank test: $V = 935, p < 0.001$). Figure 2C (left panel) plots average differences in response times for acceptances vs. rejections, against average acceptance probabilities, on the individual level. Here we see that most participants were more likely to reject than accept, and additionally rejected quickly and accepted slowly. We also observed a negative relationship between response time differences and choice probabilities, with participants who were most likely to reject the gamble also being

the ones that displayed the greatest differences in response times for acceptance vs rejection ($corr = -0.75, t(45) = 7.61, p < 0.001$ across all participants).

Both the pre-valuation bias and the prospect theory utility weighting bias can generate the response time patterns shown in Figures 2B and 2C. According to a model with only a pre-valuation bias ($\gamma < 0$) but no utility weighting bias ($\lambda = 1$), participants would need to accumulate more evidence (take longer time) when accepting rather than rejecting a gamble. According to a model with a utility weighting bias ($\lambda > 1$) but no pre-valuation bias ($\gamma = 0$), it is possible that trials on which gambles are rejected involve highly undesirable gambles (and therefore quicker response times), whereas trials on which gambles are accepted involve only moderately desirable gambles (and thus slower response times). In other words, there are two reasons as to why RT differences between acceptance and rejection might emerge: (1) participants have a pre-valuation bias towards rejecting gambles, and (2) participants reject more often in easier problems than difficult problems (Krajbich, Bartling, Hare, & Fehr, 2015).

If participants indeed have a pre-valuation bias favoring rejection, then correcting for choice factors (i.e. the gains and losses that determine the utility, and subsequently the drift rate) should not alter the qualitative relationship between choice and RT. To test this, we performed a linear regression for each participant's data, in which we regressed participants' log RTs onto gain and loss values of the gambles in the trials. We took the regression residuals as RTs adjusted for choice factors. We then grouped each participant's adjusted RTs into five bins (the first to the fifth quintiles of RTs representing the fastest to the slowest trials, respectively), and calculated each participant's rejection rate and acceptance rate for each of the adjusted RT bins, before pooling the bins for participants. The result is summarized in Figure 2D (left panel), which shows a negative relationship between choice probability and response time for rejection

decisions, and a positive relationship between choice probability and response time for acceptance decisions. This indicates that decision makers are quicker to reject and slower to accept, even when the choice factors (amounts that could be gained and lost) are controlled for. This behavioral marker is associated with a pre-valuation bias favoring rejection. As we will show in Section 3.3.3, this pattern cannot be generated by a DDM model with only the drift rate mechanisms (fixed utility bias and prospect theory utility weighting) but no pre-valuation bias.

3.3. Modeling results

3.3.1. Model fits

The model discussed in Section 2 was fit to choice and RT data using HDDM, a Python package for hierarchical Bayesian estimation of drift-diffusion models (Wiecki, Sofer, & Frank, 2013). This approach estimates group and individual level parameters simultaneously, with group-level parameters forming the prior distributions from which individual subject estimates are sampled. A recent study comparing HDDM with alternative estimation approaches showed that hierarchical fitting requires fewer data to recover parameters (Ratcliff & Childers, 2015; Wiecki et al., 2013). Moreover, the Bayesian approach permits direct inferences for parameter variability. In terms of choosing priors for the parameters, we followed the advice of Wiecki et al. (2013), and used a set of default priors that constrained parameter values in a feasible range (Matzke & Wagenmakers 2009). The complete prior specifications can be found in Wiecki et al. (2013).

We ran four separate chains for every model in all the experiments. Each chain consisted of 50,000 samples, where the first 25,000 were burn-ins and a factor of 2 was applied for thinning. To assess model convergence, we computed $\hat{R}$s of all the parameters for each model. The maximum $\hat{R}$ was 1.002 across all models, indicating successful convergence (Gelman and

Rubin, 1992). To assess model fits, we sampled 1000 sets of parameter values from their respective posterior distributions and used those to simulate participants' responses and RTs (resulting in 1000 simulated datasets, each corresponds to a set of predictions for all the trials completed by each participant). We performed posterior predictive checks by comparing posterior predictive distributions of a set of test statistics to those computed from the observed data (Figure 3, left panels). In all the presented posterior predictive checks, we took averages across all simulated datasets to compute point estimates for the test statistics, and used 95% credible intervals to quantify variabilities in predicting those test statistics. In addition, we computed correlations between the mean predicted test statistics and the observed statics across participants (Figures 3A-C) or across gambles (Figure 3D) to evaluate prediction accuracy. Our model successfully captured the observed individual heterogeneity in acceptance rates (Figure 3A; $corr = 0.99, t(47) = 45.42, p < 0.001$), and mean rejection RT (Figure 3B; $corr = 0.98, t(47) = 38.04, p < 0.001$). Some participants accepted gambles in a very small proportion of trials, and thus the correlation between observed and predicted mean acceptance RT was relatively lower (Figure 3C; $corr = 0.85, t(45) = 11.03, p < 0.001$). In addition to reproducing the relevant observed summary statistics for different participants, our model also captured the observed acceptance rates for different unique gambles (note that each of the 100 unique gambles was presented exactly twice to each participant; Figure 3D; $corr = 0.97, t(98) = 43.09, p < 0.001$). For more results on the posterior predictive tests, please see Supplemental Materials.

3.3.2. Model parameters

Table 2 summarizes the group-level parameter statistics. The posterior mean for the prospect theory utility weighting parameter ($\lambda = \frac{\beta_L}{\beta_G}$) is 1.5, and the 95% credible interval for β_L

is strictly larger than that for β_G, indicating a group-level utility weighting bias for losses in our data. The other drift rate mechanism, the fixed utility bias (α), has a posterior mean around 0. The DDM measure of pre-valuation bias (γ) has a strictly negative 95% credible interval in this experiment, indicating an overall pre-valuation bias towards rejection among our participants.

The posterior means for participant-level parameters are shown in Figure 4 (left panels). On the individual level, we observed best-fit parameter values such that $\beta_L > \beta_G$ for 39 (79.6%) participants, with 28 (57.1%) participants having a 95% credible interval for $\beta_L - \beta_G$ that is strictly positive. The posterior mean of individual-level prospect theory utility weighting parameter (λ) averaged across our participants is 1.88 (SD = 1.22; Figure 4A). We observed a negative posterior mean of α for only 24 (49.0%) participants, with a strictly negative 95% credible intervals for 9 (18.4%). Across our participants the mean value of α is -0.03 (SD = 0.49; Figure 4B). Finally, we observed a negative posterior mean of γ for 37 (75.5%) participants, 33 (67.3%) of which had a strictly negative 95% credible intervals. The averaged participant-level posterior mean of γ is -0.20 (SD = 0.24) across all participants (Figure 4C). The analyses on the individual-level parameters, together with those on the group-level, indicate that most participants display prospect theory utility weighting biases and pre-valuation biases towards rejection, but do not display a systematic fixed utility bias.

3.3.3. Constrained model analyses

To better understand the descriptive power of different mechanisms for loss aversion, we also fit three restricted variants of the DDM. The first constrained model eliminated the prospect theory utility weighting bias by setting $\beta_L = \beta_G$. The second eliminated the fixed utility bias by setting $\alpha = 0$. The third eliminated the pre-valuation bias by setting $\gamma = 0$. We compared the relative fits of these three constrained models against each other, and against the full model. The

model comparisons were performed using the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002), which measures model fits while penalizing model complexity to avoid over-fitting, with smaller DICs indicating better model fits.

The DICs of the full and constrained models are shown in Table 3. This measure revealed that despite having more parameters than the remaining models, the full model ($\text{DIC} = 17{,}184$) generated the best fit to the observed data. We used ΔDIC (differences between a constrained model and the full model) to quantify the importance of an eliminated mechanism – larger differences imply that the eliminated mechanism is more important for model fits. Out of the three constrained models, the one that set $\gamma = 0$ ($\Delta\text{DIC} = 957$) yielded the worst fit, suggesting that the pre-valuation bias plays a more important role than the other three mechanisms in this experiment. Following that was prospect theory utility weighting; the constrained model with $\beta_L = \beta_G$ led to the second largest increase in DIC ($\Delta\text{DIC} = 364$). The model without the fixed utility bias ($\alpha = 0$) yielded smaller but still notable increase in DIC ($\Delta\text{DIC} = 188$). Overall, our constrained model analyses demonstrate that all the mechanisms are indispensable in accounting for the observed data, but including the pre-valuation bias in a model explains more variance than including prospect theory utility weighting or the fixed utility bias.

3.3.4. Behavioral marker for the pre-valuation bias

Acknowledging that using goodness-of-fit as a single piece of evidence for theory testing can be problematic (Roberts & Pashler, 2000), we also used the behavioral marker shown in Figure 2D to further test for the role of the pre-valuation bias in accounting for loss aversion (note that this part of analysis was not pre-registered). This behavioral marker involves higher rejection rates for trials with shorter RTs compared to trials with longer RTs (Figure 2D, and solid blue lines in Figures 5). As discussed in Sections 2 and 3.2, this pattern is consistent with

the effect of a pre-valuation bias towards rejecting mixed gambles. To test this formally, we generated posterior predictions (1,000 samples for each trial) from the full and constrained models, and investigated how well each model mimicked the observed data. In line with our intuition, we found that the choice-RT relationship in Figure 2D can be captured by the full model, as well as by the constrained model eliminating either prospect theory utility weighting ($\beta_G = \beta_L$) or the fixed utility bias ($\alpha = 0$). However, the constrained model with no pre-valuation bias ($\gamma < 0$) fails to capture this relationship. This finding is illustrated in Figure 5 (top panels), and provides one explanation for why the pre-valuation bias plays an important role in our quantitative model fits.

3.3.5. Capturing individual heterogeneity

As we point out in the introduction section, inferences regarding utility weighting are often made by relating acceptance rates in mixed gamble tasks to psychological, clinical, and neurobiological variables. The validity of this procedure relies heavily on the assumption that utility weighting is the primary cause of loss aversion, that is, the rejection of mixed gambles. We have shown that this assumption may not hold, because an alternative mechanism, the pre-valuation bias, appears to be even more important in terms of enhancing quantitative model fits and explaining the relationship between choice and RTs. However, if the magnitude of utility weighting correlates with observable rejection rates more strongly than the pre-valuation bias, the current approach may still be acceptable.

We investigated this possibility by measuring the relationship between individual-level model parameters and observed choice heterogeneity across participants. For this purpose, we correlated participant-level estimates for λ, α and γ with average participant-level acceptance rates (Figure 6, top panel). Across participants, the acceptance rate is moderately correlated with

the fixed utility bias, α ($corr = 0.45, t(47) = 3.48, p = 0.001$), but not clearly related to utility weighting λ ($corr = -0.11, t(47) = -0.74, p = 0.462$). Instead, the correlation between acceptance rates and the pre-valuation bias, γ, is the strongest ($corr = 0.88, t(47) = 12.49, p < 0.001$). These differences in correlation persist if we use the non-parametric Spearman correlation rather than the Pearson correlation shown above (α: $corr_s = -0.28, p = 0.053$; λ: $corr_s = 0.46, p = 0.001$; γ: $corr_s = 0.90, p < 0.001$). We also performed a standardized multiple regression of acceptance rates on the parameters corresponding to the three mechanisms (Table 4), which shows that the pre-valuation bias has a considerably larger regression coefficient than prospect theory utility weighting. This indicates that a one-standard-deviation shift in the pre-valuation bias has a larger effect on the individual's gamble acceptance rates than a one-standard-deviation shift in utility weighting.

We also wanted to compare the ability of our DDM model against the standard model used in prior work, which is equipped with only utility weighting and the additive intercept. Thus we performed a standardized multiple regression to capture participant-level acceptance rates, using participant-level λ and α from the constrained DDM with the pre-valuation bias eliminated ($\gamma = 0$)[6]. The results are presented in Table 5. Comparing the R^2 of this model with the model presented in Table 4, we see that the explanatory power of the logistic-regression-comparable model (R^2=0.468) is much smaller than that of the full DDM model (R^2=0.804). This shows that participant heterogeneity predicted by the pre-valuation bias cannot be easily captured by existing models (such as a logistic model with only utility weighting and an additive intercept).

[6] Note that with a neutral pre-valuation bias ($\gamma = 0$), and a drift rate function $v_i = \alpha + \beta_G \cdot G_i - \beta_L \cdot L_i$, this constrained DDM has the same functional form for computing acceptance probabilities as the common logistic regression approach. In other words, prospect theory utility weighting and additive utility intercepts estimated from this model are comparable to those estimated from logistic regression (although the constrained DDM estimates also take into account of RT data).

Note that the models reported here do not account for the effect of risk aversion, which assumes diminishing sensitivity to monetary payoffs and is often captured by a power-transformation of objective monetary amounts. To address the concern that risk aversion may be another mechanism responsible for loss-averse behaviors, we also tested a modeling framework that incorporates risk aversion, in addition to all the mechanisms described above. All of our conclusions do generalize to that extended framework. For methods on integrating risk aversion into the DDMs, as well as results on how it (does not) affect interpretations regarding the other mechanisms, we refer the interested readers to Supplemental Materials. The Supplemental Materials also provide results of a parameter recovery analysis showing that the parameters of the model tested in the main text can be adequately recovered from data.

3.4. Summary

We used the drift diffusion model, applied to accept-reject decisions for mixed gambles, to show that loss aversion can be decomposed into multiple psychological mechanisms. Although the best-known such mechanism, prospect theory utility weighting, is indeed an important component of our model, our results show that there is also another mechanism, the pre-valuation bias, that should not be ignored. This bias has the largest quantitative contribution in terms of fitting choice and RT data. It is also necessary for capturing the choice-RT behavioral marker, and explains the most individual heterogeneity in gamble rejection rates. The concept of a pre-valuation bias has been discussed in many previous papers (e.g. Gal, 2006); here, for the first time, we demonstrate quantitatively that it is no less important than utility weighting in accounting for mixed gamble choice. By doing so we also show that prior work explaining the effects of various psychological, clinical, and neurobiological variables on loss aversion through prospect theory utility weighting may be misguided. It is more likely that these variables

influence the pre-valuation bias, given how strongly correlated this parameter is with participant-level choice outcomes.

4. Experiment 2

Given the important role of the pre-valuation bias in accounting for high rejection rates for mixed gambles, a natural follow-up question is: How do decision makers form different pre-valuation biases? How can we interpret the psychological role of such a bias? Although the pre-valuation bias resembles the status-quo bias and psychological inertia, as studied in prior work (Gal, 2006; W. Samuelson & Zeckhauser, 1988; Gal & Rucker, 2018), the DDM also offers a novel statistical interpretation: The starting point in the drift diffusion processes should reflect prior expectations regarding how rewarding different responses are. Consistent with this interpretation, empirical evidence in perceptual decision making studies finds that the starting point can be shifted when participants' expectations change. For example, if participants discover that a certain response is more likely to be the more rewarding response than the other, they adapt to this knowledge and become more prepared to choose the more rewarding response by shifting their starting point towards the threshold corresponding to that response.

The objective of Experiment 2 was to test whether people's expectations about mixed gambles have a similar influence on the pre-valuation bias and thus on loss aversion. We did this by altering the set of gambles shown to participants in a design similar to that of Experiment 1. In one condition of Experiment 2, the gambles were mostly undesirable, and we expected participants to display a negative pre-valuation bias, corresponding to a prior expectation favoring rejection. In another condition, the gambles were mostly desirable, and we expected participants to display a positive (or, at the very least, less negative) pre-valuation bias,

corresponding to a prior expectation favoring acceptance. If pre-valuation biases can indeed be shifted by altering the desirability of previously encountered gambles, we could interpret these biases to be adaptive to prior beliefs about the gamble desirability. We preregistered our experimental methods, model specifications and main hypotheses at https://osf.io/varx6.

4.1 Methods

101 participants (age: mean = 25.96, SD = 9.75; 56.4% female) were recruited from a university experimental research subject pool and performed the experiment on computers in a laboratory. Each participant followed the same procedure as in Experiment 1, i.e., they made 200 decisions to accept or reject gambles involving equal probabilities of gains and losses. As in Experiment 1, we recorded their choice and RT data, and incentivized participants by randomly selecting one of the trials to play out.

Unlike Experiment 1, participants in Experiment 2 were assigned to one of two conditions. In the high-payoff (HP) condition, the possible gain values in a gamble were taken from the set of {60, 70, 80, 90, 100, 110, 120, 130, 140, 150} tokens; whereas the possible loss values were taken from the set of {-10, -20, -30, -40, -50, -60, -70, -80, -90, -100} tokens. In the low-payoff (LP) condition, the possible gain values in a gamble were taken from the set of {10, 20, 30, 40, 50, 60, 70, 80, 90, 100} tokens; whereas the possible loss values were taken from the set of {-60, -70, -80, -90, -100, -110, -120, -130, -140, -150} tokens. With presentation positions counterbalanced, we had in a total of 2 × 10 × 10 = 200 trials for both conditions. The distributions of values of the gambles overlapped between the conditions, so that participants in both conditions completed gambles with gains from the set of {60, 70, 80, 90, 100} tokens, and losses from the set of {-60, -70, -80, -90, -100} tokens. With position counterbalanced, this

resulted in a total of 2 × 5 × 5 = 50 shared gambles (see Figure 1C for a visual illustration of the payoff distribution design).

As in Experiment 1, each token was worth US$0.10. However, since the LP participants could potentially have a larger loss, all participants in Experiment 2 began the experiment with an endowment of 150 tokens (US$15). There were 52 participants in the HP condition (age: mean = 26.65, SD = 9.67; 59.6% female), and 49 participants in the LP condition (age: mean = 25.22, SD = 9.87; 53.1% female).

If our hypotheses regarding the causes of the pre-valuation bias are correct, then we would expect choices in the HP condition to display higher pre-valuation biases (corresponding to acceptance being more rewarding) and choices in the LP condition to display lower pre-valuation biases (corresponding to rejection being more rewarding). This would indicate that participants form pre-valuation biases based on (learnt) expectations regarding the gambles. If the pre-valuation biases are identical across the two conditions then it is likely that pre-valuation biases are not a product of environment adaptivity, and are instead caused by some other (potentially non-malleable) component of decision processes.

Note that because HP and LP participants experienced different payoff distributions, the relative attractiveness of the gambles could also be different for the participants in the two conditions. For example, for the HP participants, the shared gambles had the lowest gain values and the highest loss values, and thus were the worst gambles available. The reverse was true for LP participants, for whom the shared gambles were the best available. This desirability difference could alter the drift rate, if the drift rate depends on relative attractiveness (see Walasek & Stewart, 2015 for a related account of this phenomena). Importantly, this effect is different from the pre-valuation bias effect that is the primary focus of our experiment.

4.2. Descriptive Results

4.2.1. Overview of the whole dataset

In the HP condition, the mean rejection rate was 50.9% (SD = 0.20), with 51.9% of participants rejecting more than half of the gambles. In the LP condition, the mean rejection rate was 88.9% (SD = 0.11), with all participants rejecting more than half of the gambles. The difference in overall rejection rates is significant ($\Delta\text{Mean} = 38.1\%, t(78) = 11.92, p < 0.001$). The large acceptance rate discrepancy between the two conditions demonstrates the success of our payoff distribution manipulation.

In both conditions, participants tended to accept the gambles when the size of the gain exceeded the size of the loss. For HP participants, the size of the gain had to be more than 1.83 times the size of the loss; whereas for LP participants the size of the gain only had to be more than 1.25 time the size of the loss (Figure 2A, middle and right panels). In the HP condition, the average rejection decision took 1.54 seconds, whereas the average acceptance decision took 1.59 seconds. This difference is not statistically significant (Figure 2B, middle panel; $t(51) = 0.63, p = 0.529$). In the LP condition, the average rejection decision took 1.26 seconds and the average acceptance decision took 2.50 seconds. This is a significant difference (Figure 2B, right panel; $t(43) = 7.33, p < 0.001$). Compared to participants in the LP condition, participants in the HP condition were quicker when accepting gambles ($\Delta\text{Mean} = -0.91$ seconds, $t(67) = -4.81$, $p < 0.001$), and slower when rejecting gambles ($\Delta\text{Mean} = 0.28$ seconds, $t(73) = 2.52$, $p = 0.014$). Consistent with Experiment 1, participants who were more likely to reject the gamble also displayed larger differences in response times for acceptance vs rejection (Figure 2C, middle and right panels; HP: $corr = -0.85, t(50) = 11.56, p < 0.001$; LP: $corr = -0.58, t(42) = 4.57, p < 0.001$).

Finally, we also tested whether the behavioral marker identified in Experiment 1 surfaced in Experiment 2. To do so, we plotted the acceptance and rejection rates against adjusted RT bins (Figure 2D, middle and right panels). As in Experiment 1, participants in both conditions were quicker to reject than accept gambles, when choice factors (amounts to be gained and lost) were controlled for. Although the acceptance rates were higher in the HP condition and lower in the LP condition, there was a negative relationship between RT and acceptance rates in both conditions.

Our descriptive analyses demonstrate that the findings in Experiment 1 are generalizable to a similar mixed gamble experiment with distinct payoff distributions. They also indicate that our hypotheses regarding pre-valuation biases may be correct: Participants in the LP condition were much quicker to reject than accept than in the HP condition, indicating that they may have had pre-valuation biases closer to the rejection boundary. Additionally, participants in the LP condition had higher rejection probabilities. This could be caused by more negative pre-valuation biases, though it may also be the case that the two conditions generated differences in drift rates.

4.2.2. Examining the shared gambles

Participants in both conditions completed gambles with gain values from the set {60, 70, 80, 90, 100} tokens, and loss values from the set {-60, -70, -80, -90, -100}. We refer to these 25 gambles (50 trials with position counterbalanced) as shared gambles. In this section we analyze behavioral patterns for shared gambles. Note that this analysis was not preregistered.

Although HP participants accepted more gambles than LP participants, their acceptance rates were lower than the LP participants if we only analyze the shared gambles (HP: 13.3%; LP: 34.6%; $t(81) = 4.86, p < .001$). On the other hand, the RT patterns emerged from the shared gambles were more consistent with the whole dataset. HP participants were quicker in

acceptance decisions than LP participants (HP: 1.81; LP: 2.44; $t(66) = 2.15, p = 0.035$). The difference in rejection RTs was nonsignificant (HP: 1.45; LP: 1.50; $t(94) = 0.43, p = 0.671$). We generated a composite variable to describe RT patterns for participants that at least accepted or rejected in one trial. The composite variable is referred to as the averaged RT difference, and is defined as the difference between mean acceptance RTs and mean rejection RTs for a participant. HP participants exhibited smaller averaged RT differences than LP participants (HP: 0.19; LP: 0.88; $t(76) = 2.77, p = 0.007$). Note that the fact that HP participants accepted fewer gambles but took relatively less time to accept, contradicts the usual choice-RT correlations observed in binary choice data. In the following section we will get back to this result and explain it using the DDM decomposition.

4.3. Modeling results

4.3.1. Replication of Experiment 1

To rigorously test for differences in parameters across the two conditions, and to more formally replicate Experiment 1, we fit the DDM to the HP and LP data separately. As in Experiment 1, our posterior predictive checks showed that our full model fits successfully captured the observed heterogeneity in acceptance rates (Figure 3A, middle and right panels), mean rejection RT (Figure 3B, middle and right panels), and mean acceptance RTs (Figure 3C, middle and right panels) across participants, as well as the observed heterogeneity in acceptance rates across gambles (Figure 3D, middle and right panels), for the two conditions.

We also examined the role of different DDM mechanisms using constrained model analyses. Table 3 shows model fits for the full and constrained models in both conditions. Here, the full models include the pre-valuation bias (γ), the utility weighting bias (λ), and the fixed utility bias (α). In the HP condition, unlike Experiment 1, eliminating utility weighting (ΔDIC =

1,071) produced a larger DIC increase than eliminating the pre-valuation bias ($\Delta\text{DIC} = 769$). Similar to the findings of Experiment 1, eliminating the fixed utility bias resulted in a smaller yet still noticeable DIC increase ($\Delta\text{DIC} = 362$). In the LP condition, eliminating the pre-valuation bias increased DIC the most ($\Delta\text{DIC} = 914$). Compared to this, eliminating the remaining two mechanisms had a much smaller effect (utility weighting bias: $\Delta\text{DIC} = 186$; fixed utility bias: $\Delta\text{DIC} = 129$). Thus, replicating what we found in Experiment 1, the three DDM mechanisms are all indispensable when accounting for the variance in choice and RT data; eliminating any of them worsens model fits. However, the relative importance of the pre-valuation bias and utility weighting bias varies across the HP and LP conditions. As we will show in Section 4.3.2, this is because participants displayed different magnitudes of pre-valuation biases and utility weighting biases in the two experimental conditions.

As in Experiment 1, we used posterior predictive samples to test which DDM mechanism(s) were essential in accounting for the behavioral marker in Figure 2D, i.e., higher rejection rates in quicker trials with choice factors controlled for. As shown in Figure 5 (middle and right panels), the choice-RT behavioral marker can be captured by the full model, as well as by the constrained models eliminating the utility weighting and fixed utility biases. However, the constrained model with no pre-valuation bias fails to capture this relationship. In other words, the pre-valuation bias is necessary for explaining the choice-RT pattern. This is the case in both the HP and LP conditions.

Our final replication attempt involves predicting participant-level acceptance rates using the three DDM mechanisms through standardized regressions. The results are shown in Table 4. As before, the coefficients associated with the pre-valuation bias were higher than those for the other mechanisms. This was the case in both conditions. This result can also be seen in Figure 6

(middle and right panels), in which the pre-valuation bias has the highest correlation with participant-level acceptance rates.

4.3.2 Model parameters

Now we examine differences across the two conditions. We begin by comparing group-level and individual-level parameters, which are summarized in Table 2 and Figure 4 respectively. These show that participants in the LP condition exhibited more negative pre-valuation biases than their counterparts in the HP condition. On the group level, the posterior mean of the pre-valuation bias (γ) is -0.07 (95%CI $[-0.11, -0.02]$) in the HP condition, and -0.32 (95%CI [-0.35, -0.28]) in the LP condition. On the individual level, in the HP condition, 33 (63.5%) participants had a negative posterior mean for the pre-valuation bias, with 24 (46.2%) of them exhibiting a strictly negative 95% CI for the parameter. In the LP condition, 49 (100%) had a negative posterior mean for the pre-valuation bias, out of which 47 (95.9%) had a strictly negative 95% CI. Overall, the difference between the biases in the two conditions was very large (HP: Mean $= -0.06$, SD $= 0.22$; LP: Mean $= -0.31$, SD $= 0.13, t(84) = 7.05, p < 0.001$; based on posterior means of parameters). Therefore, in line with our predictions, manipulating payoff distributions across conditions caused shifts in participants' pre-valuation biases.

Although the experiment was set up to test the influence of payoff distributions on the pre-valuation bias, we also predicted (based on Walasek & Stewart, 2015) that utility weighting may depend on the pay-off distribution. We found that this was in fact true in our data. Although group-level posterior distributions of β_G were similar between the two conditions (HP: Mean $=$ 0.016; LP: Mean $= 0.016$), LP participants had a much lower β_L (HP: Mean $=$ 0.029; LP: Mean $= 0.013$) and much more negative fixed utility biases in the drift rate (HP: Mean $= 0.014$;

LP: Mean $= -0.420$) than participants in the HP condition. This can also be seen on the individual level, HP participants had stronger utility weighting biases with larger values of λ (HP: Mean $= 2.11$, SD $= 1.36$; LP: Mean $= 1.03$, SD $= 0.50, t(65) = 5.30, p < 0.001$), but they exhibited a less positive fixed utility bias than LP participants (HP: Mean $= 0.01$, SD $= 0.81$; LP: Mean $= -0.42$, SD $= 0.36, t(71) = 3.49, p < 0.001$). In the HP condition, 49 (94.2%) participants exhibited a utility weighting bias ($\lambda > 1$), but only 28 (53.8%) had a negative fixed utility bias ($\alpha < 0$); whereas in the LP condition, only 20 (40.8%) had $\lambda > 1$, but 43 (87.8%) had $\alpha < 0$. The differences in λ emerged because HP and LP participants had very similar β_G (HP: Mean $= 0.016$, SD $= 0.006$; LP: Mean $= 0.016$, SD $= 0.009, t(82) = 0.16, p = 0.872$), but HP participants had a much lager β_L (HP: Mean $= 0.029$, SD $= 0.011$ LP: Mean $= 0.013$, SD $= 0.005, t(75) = 9.42, p < 0.001$).

4.3.3. Shared gambles

The differences in λ documented above suggest that, for any individual gamble, the disutility caused by the loss is smaller for LP participants than HP participants. Therefore, a gamble with a large loss may appear better to LP participants than their HP counterparts. To test this, we examined the drift rates of the fifty gambles that were shared across the conditions. Specifically, with the individual-level parameters estimated based on the whole dataset, we reconstructed the mean drift rate ($\bar{v}$) of the shared gambles for each participant. As hypothesized, we found that HP participants had more negative mean drift rates for the shard gambles than LP participants (HP: Mean $= -1.06$, SD $= 0.65$; LP: Mean $= -0.20$, SD $= 0.51, t(96) = 7.39, p < 0.001$). This indicates that HP participants considered the shared gambles to be very unappealing; whereas LP participants were almost neutral between accepting vs. rejecting these gambles. The difference in utility for shared gambles is likely due to the discrepancy in relative

gamble attractiveness in the two conditions: the shared gambles offered smaller gains (losses) and larger losses (gains) to HP (LP) participants compared to the other non-shared gambles.

The specific loss and gain values in a gamble should not influence decision makers' pre-valuation bias as, by definition, this bias captures a predisposition prior to utility evaluation. To verify whether our pre-valuation bias estimates were truly gamble-insensitive, we fit the full DDM to shared gambles only, for the two groups. The pre-valuation biases recovered from shared gambles were very similar to those estimated from the whole dataset (HP: $corr = 0.86, t(50) = 11.84, \mathrm{p} < 0.001$; LP: $corr = 0.76, t(47) = 8.09, \mathrm{p} < 0.001$).

To summarize, compared to LP participants, HP participants had a more positive pre-valuation bias, and yet simultaneously a more negative combined drift rate. This is shown in Figure 7A which displays a scatter plot of participant-level mean drift rates for the shared gambles and pre-valuation biases (using parameters inferred from choices on all the gambles). By observing the two parameters changing in the opposite directions for the shared gambles, we demonstrate a dissociation between drift rate and the pre-valuation bias mechanisms.

This disassociation also yields some interesting behavioral patterns in our data. As shown in Figure 7B, the shared gamble acceptance rates were on average lower in the HP condition (HP: Mean $= 13.3\%$, SD $= 16.7\%$; LP: Mean $= 34.6\%$, SD $= 26.0\%$). However, at the same time, HP participants also exhibited smaller differences between acceptance RTs and rejection RTs than LP participants (HP: Mean $= 0.19$, SD $= 1.13$; LP: Mean $= 0.88$, SD $= 1.11$). The fact that HP participants accepted fewer gambles but took relatively less time to accept, contradicts the usual choice-RT correlations observed in binary choice data. This odd pattern however stems naturally from the disassociation we observed between the pre-valuation bias and the drift rate mechanisms. If choice rates depend primarily on drift rates, then we would expect

HP participants to have lower acceptance probabilities than LP participants. Likewise, if RT differences depend primarily on pre-valuation biases, then we would expect LP participants to have larger differences in acceptance and rejection RTs than HP participants. Indeed, our shared gamble data shows that the acceptance rate is positively correlated with the combined drift rate (Figure 7C; $corr = 0.86, t(99) = 16.69, p < 0.001$) on the participant-level when pooling the HP and LP participants. On the contrary, the correlation between acceptance rates and pre-valuation biases is much weaker (Figure 7D; $corr = 0.26, t(99) = 2.65, p = 0.009$). Likewise we find that the averaged RT difference between acceptance vs. rejection decisions is highly correlated with the pre-valuation bias across all participants (Figure 7F; $corr = -0.48, t(79) = -4.80, p < 0.001$), but there is no overall correlation between the averaged RT difference and the combined drift rate (Figure 7E; $corr = -0.07, t(79) = -0.64, p = 0.523$).

4.4. Summary

The results of Experiment 2 replicate Experiment 1, and show that the pre-valuation bias has an important role in quantitative model fits, in describing choice-RT relationships and accounting for the behavioral marker, and in explaining gamble acceptance rates across participants. In addition, they also show that the pre-valuation bias can be manipulated by changing the gambles that participants are exposed to. If participants are given primarily undesirable gambles (as in the LP condition), they form priors expectations favoring gamble rejection and display more negative pre-valuation biases, whereas if they are given more desirable gambles (as in the HP condition) they form relatively neutral prior expectations and display less negative pre-valuation biases. Thus, the pre-valuation bias can be interpreted as an adaptive consequence of expectations regarding gamble desirability. Its importance in explaining behavioral data suggests that these adaptive beliefs play a major role in risky choice.

Experiment 2 also documented a disassociation between the pre-valuation bias and the drift rate (utility weighting and fixed utility bias) mechanisms, which generated some interesting behavioral patterns for gambles shared across the two conditions. As predicted, participants were more likely to have pre-valuation biases favoring rejection when exposed to undesirable gambles (LP condition) than when exposed to desirable gambles (HP condition). But for gambles shared by the two conditions, they showed a reduced utility weighting bias (lower λ), and thus a higher drift rate, in the LP condition than the HP condition. This reduces the rejection rates, while also increasing the RT for acceptance (and reducing the RT for rejection) for the shared gambles in the LP condition relative to the shared gambles in the HP condition. The higher utility (drift rate) for the shared gambles in the presence of other undesirable gambles can be easily explained if participants base their utilities on relative comparisons with prior gambles, as has been shown by Walasek and Stewart (2015). In this sense, Experiment 2 can be seen as a conceptual replication of the findings of this earlier paper.

5. Discussion

Understanding why people are loss averse is a central goal for researchers interested in theory development and in real world applications of risky choice. The predominant explanation for this phenomenon is prospect theory's utility weighting - the larger effect of losses vs. gains on gamble utility. In fact, a lot of prior work in psychology, economics, and neuroscience infers utility weighting through mixed gamble rejection rates, and subsequently uses this explanation of loss aversion to characterize the effect of social, cognitive, emotional, developmental, demographic, clinical, physiological, and neural variables on risky choice. This fundamental assumption shared by these studies has been challenged over the recent years. Some researchers

argue that a fixed utility bias should also be included in gamble valuation; others argue that decision makers reject gambles because of the status quo bias.

In this paper, we revisit the concept of loss aversion, viewing it as a decision tendency that emerges from an interplay of multiple unique psychological mechanisms. Using a drift-diffusion decomposition of mixed gamble accept-reject decisions, we precisely define and quantitatively measure these mechanisms simultaneously on an individual level, and subsequently evaluate their relative importance. We find that the utility weighting bias, the fixed utility bias, and the pre-valuation bias, are all crucial for explaining observed variance in choice and RT data, but that the pre-valuation bias has the largest effect on model fits. Additionally, this bias is associated with a unique behavioral marker: faster RTs for rejection than acceptance (controlling for various choice factors). Our experiments show strong evidence for this behavioral marker, indicating that the pre-valuation bias is essential for capturing the relationship between choices and response times. This pre-valuation bias is also more correlated with participant-level acceptance rates than other loss aversion mechanisms in the risky choice task.

In Experiment 2, we further demonstrate that the pre-valuation bias depends on participants' expectations about gamble desirability, so that manipulating the gambles that decision makers are exposed to can alter the strength of this bias. These results suggest that the pre-valuation bias stems from biased prior expectations regarding different responses, and thus forms part of an adaptive strategy for making risky choices with the potential of loss. We also dissociate the effects of the pre-valuation bias and the drift rate mechanisms. In this study, we evaluate people's choices on a set of shared gambles having exposed different groups of participants to unique sets of either high-payoff (HP) or low-payoff (LP) gambles. Because the pre-valuation bias depends on prior experience, and is insensitive to specific gamble values in a

given trial, HP participants have more positive pre-valuation biases than LP participants when deliberating over the shared gambles. However, because the relative desirability of the shared gambles differs between the groups, HP participants exhibit higher utility weighting biases and more negative drift rates than their LP counterparts. Importantly, the divergence between these mechanisms leads to interesting behavioral effects: HP participants are less likely to accept shared gambles but do so relatively quickly compared to LP participants. This dissociation in both parameter estimates and behavioral patterns demonstrates that loss aversion is the product of distinct pre-valuation and utility valuation mechanisms.

Our results have important implications for how we interpret people's tendency to reject mixed gambles. Given that multiple mechanisms cause loss aversion, attributing heterogeneity in acceptance rates to utility weighting alone may cause serious reverse inference problems: Variables originally attributed to utility weighting may be better understood in terms of pre-valuation bias tendencies (especially as the latter correlate more strongly with individual-level acceptance rates). For example, the well-known finding that ventral striatum activity correlates with mixed gamble rejection rates (Tom et al., 2007) could be due to the relationship between brain activity and the pre-valuation bias rather than the relationship between brain activity and utility weighting, as is commonly assumed. Additional research is needed to untangle these relationships and future work should consider the possibility that gamble rejection rates, as well as the psychological, clinical and neurobiological correlates of high rejection rates, can be understood in terms of multiple different psychological mechanisms (including those at play prior to the formation of utility).

The findings of this paper also add theoretical nuance to some recently documented empirical results regarding loss aversion. Walasek and Stewart (2015), for example, find that loss

aversion can be altered by altering the distribution of gambles shown to participants. More recently, Walasek and Stewart (2019) showed that such effects can be captured by the composite measure of the additive bias (in the logistic regression) and asymmetric weighting of gains and losses. Our findings that the HP and LP conditions in Experiment 2 differ in terms of their drift rates identifies the psychological mechanism responsible for this phenomenon: It is gamble valuation, and not the pre-valuation response tendency, that is affected by relative comparisons with previously experienced gambles. Additionally, by showing that the pre-valuation bias is an important mechanism underlying loss aversion, we are able to explain why loss aversion diminishes when the accept-reject task is reframed as a choice between the gamble and a sure outcome of zero (Ert & Erev, 2008, 2013): The choice task does not involve an explicit status quo, and thus participants are less likely to develop a pre-valuation bias. In a preliminary study (not reported in the current paper) we have confirmed this hypothesis by explicitly fitting the proposed DDM model to mixed gamble choices. In future work we plan to build off these results in order to provide a more detailed formal account of the contextual determinants of loss aversion.

Our study of the pre-valuation bias also raises many new research questions. For example, we have shown that the pre-valuation bias adapts to prior experience, with a less rewarding environment predisposing participants towards rejection decisions and a more rewarding environment disposing participants towards acceptance decisions. However, there are important asymmetries in this process. For example, Experiment 1 offers gambles that have, on average, equal sized gains and losses, and thus an average expected value of zero. Yet it finds that there is an overall pre-valuation bias for rejection. Similarly, in Experiment 2 we find that participants are reluctant to shift the pre-valuation bias towards the acceptance boundary. Even

though participants in the high payoff condition received highly attractive gambles with large positive expected values, they still displayed a mild pre-valuation tendency to reject.

Does this resistance to positive pre-valuation biases reflect tendencies to reject gambles established outside of the lab? It seems so. We performed additional analyses using the first 25 gambles completed by each participant (which constitute the first half of the first experimental block). With limited within-experiment experience, pre-valuation biases should be reflective of participants' life experiences and pre-experimental response tendencies. We found that the behavioral marker for the pre-valuation bias was also observable at the very beginning of the experiments (Figure 8), indicating that participants might bring a negative pre-valuation bias into the experiment. To better understand why this happens researchers need to develop models that specify the relationship between risky choice and long-term experience in everyday decision environments. There have been some attempts at this (see e.g. Leuker et al., 2018; Pleskac & Hertwig, 2014; Stewart et al., 2006; Stewart, Reimers & Harris, 2015), and future work should apply the insights of this research program to the study of the psychological mechanisms underpinning loss aversion.

The predominant approach in modeling risky choice involves computing subjective expected utility of gambles. Our paper suggests that using a dynamic cognitive model can enable us to make interpretations regarding a fuller set of psychological mechanisms underlying these decisions. Going from interpretations to predictions, risky models that take into account multiple cognitive processes also make better out-of-sample predictions for human choice, compared to models based on computing subjective expected utilities (e.g., Erev, Ert, Plonsky et al., 2017). Our work therefore also contributes to the growing body of research showing the role of information processing in interpreting various parameters of descriptive models of choice

(Ashby, Yechiam, & Ben-Eliezer, 2018; Pachur, Schulte-Mecklenbeck, Murphy, & Hertwig, 2018).

The tests presented in this paper rely on drift diffusion model fits to choice and RT data. Without this modeling framework we would not be able to identify and measure the pre-valuation bias. The validity of the drift diffusion model (including the importance of the starting point bias) has been shown in prior experiments involving perceptual, lexical, motor tasks (Forstmann, Ratcliff, & Wagenmakers, 2016; Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Ratcliff & Rouder, 2000; Ratcliff, Gomez, & McKoon, 2004; Ratcliff, Smith, Brown, & McKoon, 2016; White & Poldrack, 2014), and preferential choice tasks (Bhatia, 2014; Busemeyer & Townsend, 1993; Clay et al., 2017; Dai & Busemeyer, 2014; Krajbich, Armel, & Rangel, 2010; Philiastides & Ratcliff, 2013; Trueblood, et al., 2014; Tsetsos et al., 2016; Turner, et al., 2018; Zhao et al., 2019). We further illustrate the value of this approach for understanding the psychological processes at play in acceptance-rejection decisions for a mixed gamble task. The dynamic modeling approach has also been used to study other risk preferences, including binary choices between gambles (e.g., Diederich & Trueblood, 2018), and even riskless choices involving the integration of gain and loss attributes (e.g., Horn, Mata & Pachur, 2020). The mechanisms decomposed from such models can be connected to gaze allocation, pupil dilation, and brain activities (e.g., Basten, Biele, Heekeren & Fiebach, 2010; Shengn, Ramakrishnan, Seok, et al., 2020; Turner, Forstmann & Steyvers, 2019). These results suggest that the DDM, and related dynamic models, offer a cohesive general framework, one that can describe how decisions begin, how they unfold over time, how they are terminated, and how they can be separated into several measurable and interpretable mechanisms. Such models are necessary to bridge the gap between formal cognitive theory and observable behavioral and physiological

outcomes with real world consequence, and we look forward to their continued application to the study of human decision making.

References

Ashby, N. J., Yechiam, E., & Ben-Eliezer, D. (2018). The consistency of visual attention to losses and loss sensitivity across valuation and choice. *Journal of Experimental Psychology: General*, 147(12), 1791.

Barkley-Levenson, E. E., Van Leijenhorst, L., & Galván, A. (2013). Behavioral and neural correlates of loss aversion and risk avoidance in adolescents and adults. *Accident Analysis and Prevention*, *3*, 72–83.

Barkley-Levenson, E., & Galvan, A. (2014). Neural representation of expected value in the adolescent brain, *111*(4), 1646–1651.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Science*s, 107(50), 21767-21772.

Benartzi, S., & Thaler, R. H. (1999). Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments. *Management Science*, *45*(3), 364–381.

Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review*, *21*(5), 1095–1111.

Bhatia, S. (2017). Comparing theories of reference-dependent choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1490–1507.

Bibby, P. A., & Ferguson, E. (2011). The ability to process emotional information predicts loss aversion. *Personality and Individual Differences*, *51*(3), 263–266.

Birnbaum, M. H. (2008). New paradoxes of risky decision making. *Psychological Review*, *115*(2), 463–501.

Bogacz, R. (2007). Optimal decision-making theories: linking neurobiology with behaviour. *Trends in Cognitive Sciences*, *11*(3), 118–125.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment. *Psychological Review*, *100*(3), 432–459.

Canessa, N., Crespi, C., Baud-Bovy, G., Dodich, A., Falini, A., Antonellis, G., & Cappa, S. F. (2017). Neural markers of loss aversion in resting-state brain activity. *NeuroImage*, *146*(C), 257–265.

Canessa, N., Crespi, C., Motterlini, M., Baud-Bovy, G., Chierchia, G., Pantaleo, G., et al. (2013). The Functional and Structural Neural Basis of Individual Differences in Loss Aversion. *Journal of Neuroscience*, *33*(36), 14307–14317.

Clay, S. N., Clithero, J. A., Harris, A. M., & Reed, C. L. (2017). Loss aversion reflects information accumulation, not bias: a drift-diffusion model study. *Frontiers in psychology*, 8, 1708.

Clithero, J. A. (2018). Improving out-of-sample predictions using response times and a model of the decision process. *Journal of Economic Behavior & Organization*, *148*, 344–375.

Dai, J., & Busemeyer, J. R. (2014). A probabilistic, dynamic, and attribute-wise model of intertemporal choice. Journal of Experimental Psychology: General, 143(4), 1489.

De Martino, B., Camerer, C. F., & Adolphs, R. (2010). Amygdala damage eliminates monetary loss aversion, *107*(8), 3788–3792.

Diederich, A., & Trueblood, J. S. (2018). A dynamic dual process model of risky decision making. *Psychological review*, 125(2), 270.

Engelmann, J. B., Meyer, F., Fehr, E., & Ruff, C. C. (2015). Anticipatory Anxiety Disrupts Neural Valuation during Risky Choice. *Journal of Neuroscience*, *35*(7), 3085–3099.

Ert, E., & Erev, I. (2008). The rejection of attractive gambles, loss aversion, and the lemon avoidance heuristic. *Journal of Economic Psychology*, 29, 715–723.

Ert, E., & Erev, I. (2013). On the Descriptive Value of Loss Aversion in Decisions under Risk. *Judgment and Decision Making*, *8*(3), 214–235.

Erev, I., Ert, E., Plonsky, O., Cohen, D., & Cohen, O. (2017). From anomalies to forecasts: Toward a descriptive model of decisions under risk, under ambiguity, and from experience. Psychological review, 124(4), 369.

Erev, I., Ert, E., & Yechiam, E. (2008). Loss aversion, diminishing sensitivity, and the effect of experience on repeated decisions. *Journal of Behavioral Decision Making*, *21*(5), 575–597.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions. *Annual Review of Psychology*, *67*(1), 641–666.

Gal, D. (2006). A psychological law of inertia and the illusion of loss aversion. *Judgment and Decision Making*, *1*(1), 23–32.

Gal, D., & Rucker, D. D. (2018). The Loss of Loss Aversion: Will It Loom Larger Than Its Gain? *Journal of Consumer Psychology*, *28*(3), 497–516.

Gelskov, S. V., Henningsson, S., Madsen, K. H., Siebner, H. R., & Ramsøy, T. Z. (2015). Amygdala signals subjective appetitiveness and aversiveness of mixed gambles. *Cortex*, *66*(C), 81–90.

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574.

Hadlaczky, G., Hökby, S., Mkrtchian, A., Wasserman, D., Balazs, J., Machín, N., et al. (2018). Decision-Making in Suicidal Behavior: The Protective Role of Loss Aversion. *Frontiers in Psychiatry*, *9*, e1000123–9

Harinck, F., Van Dijk, E., Van Beest, I., & Mersmann, P. (2007). When gains loom larger than losses: reversed loss aversion for small amounts of money. *Psychological Science*, *18*(12), 1099–1105.

Horn, S. S., Mata, R., & Pachur, T. (2020). Good+ Bad=? Developmental Differences in Balancing Gains and Losses in Value-Based Decisions From Memory. Child development, 91(2), 417-438.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697–720.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263.

Kermer, D. A., Driver-Linn, E., Wilson, T. D., & Gilbert, D. T. (2006). Loss Aversion Is an Affective Forecasting Error. *Psychological Science*, *17*(8), 649–653.

Kőszegi, B., & Rabin, M. (2007). Reference-Dependent Risk Attitudes. *American Economic Review*, *97*(4), 1047–1073.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*(1), 7455.

Lazzaro, S. C., Rutledge, R. B., Burghart, D. R., & Glimcher, P. W. (2016). The Impact of Menstrual Cycle Phase on Economic Choice and Rationality. *PloS One*, *11*(1), e0144080–15.

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making*, *6*(7), 651–687.

Leuker, C., Pachur, T., Hertwig, R., & Pleskac, T. J. (2018). Exploiting risk–reward structures in decision making under uncertainty. *Cognition*, 175, 186-200.

Lorains, F. K., Dowling, N. A., Enticott, P. G., Bradshaw, J. L., Trueblood, J. S., & Stout, J. C. (2014). Strategic and non-strategic problem gamblers differ on decision-making under risk and ambiguity. *Addiction*, *109*(7), 1128–1137.

Markett, S., Heeren, G., Montag, C., Weber, B., & Reuter, M. (2016). Loss aversion is associated with bilateral insula volume. A voxel based morphometry study. *Neuroscience Letters*, *619*, 172–176.

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *32*(7), 2335–2343.

Pachur, T., & Kellen, D. (2013). Modeling gain-loss asymmetries in risky choice: The critical role of probability weighting. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

Pachur, T., & Scheibehenne, B. (2017). Unpacking buyer-seller differences in valuation from experience: A cognitive modeling approach. *Psychonomic bulletin & review*, 24(6), 1742-1773.

Pachur, T., Schulte-Mecklenbeck, M., Murphy, R. O., & Hertwig, R. (2018). Prospect theory reflects selective allocation of attention. *Journal of experimental psychology: general*, 147(2), 147.

Pighin, S., Bonini, N., Savadori, L., Hadjichristidis, C., & Schena, F. (2014). Loss aversion and hypoxia: less loss aversion in oxygen-depleted environment. *Stress*, *17*(2), 204–210.

Pleskac, T. J., Wallsten, T. S., Wang, P., & Lejuez, C. W. (2008). Development of an automatic response mode to improve the clinical utility of sequential risk-taking tasks. *Experimental and clinical psychopharmacology*, 16(6), 555.

Polman, E. (2012). Self–other decision making and loss aversion. *Organizational Behavior and Human Decision Processes*, *119*(2), 141–150.

Rabin, M. (2000). Diminishing marginal utility of wealth cannot explain risk aversion. In D. Kahneman & A. Tversky (Eds.), *Choices, Values and Frames*. New York.

Rabin, M., & Thaler, R. H. (2001). Anomalies: Risk Aversion. *Journal of Economic Perspectives*, *15*(1), 219–232.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Ratcliff, R., & Childers, R. (2015). Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making. *Decision*, *2015*.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, *111*(1), 159–182.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 127–140.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. doi:10.1037/0033-295x.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281.

Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1446.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Samuelson, P. A. (1960). The St. Petersburg Paradox as a Divergent Double Limit. *International Economic Review*, *1*(1), 31–37.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59.

Schulreich, S., Gerhardt, H., & Heekeren, H. R. (2016). Incidental fear cues increase monetary loss aversion. *Emotion*, *16*(3), 402–412.

Sheng, F., Ramakrishnan, A., Seok, D., Zhao, W. J., Thelaus, S., Cen, P., & Platt, M. L. (In Press). Decomposing loss aversion from gaze allocation and pupil dilation. *Proceedings of the National Academy of Sciences*.

Simonson, I., & Kivetz, R. (2018). Bringing (Contingent) Loss Aversion Down to Earth — A Comment on Gal & Rucker's Rejection of ‘Losses Loom Larger Than Gains’. *Journal of Consumer Psychology*, *28*(3), 517–522.

Sokol-Hessner, P., Camerer, C. F., & Phelps, E. A. (2013). Emotion regulation reduces loss aversion and decreases amygdala responses to losses. *Social Cognitive and Affective Neuroscience*, *8*(3), 341–350.

Sokol-Hessner, P., Hartley, C. A., Hamilton, J. R., & Phelps, E. A. (2015a). Interoceptive ability predicts aversion to losses. *Cognition & Emotion*, *29*(4), 695–701.

Sokol-Hessner, P., Lackovic, S. F., Tobe, R. H., Camerer, C. F., Leventhal, B. L., & Phelps, E. A. (2015b). Determinants of Propranolol's Selective Effect on Loss Aversion. *Psychological Science*, *26*(7), 1123–1130.

Sokol-Hessner, P., Raio, C. M., Gottesman, S. P., Lackovic, S. F., & Phelps, E. A. (2016). Acute stress does not affect risky monetary decision-making. *Neurobiology of Stress*, *5*(C), 19–25.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639.

Stewart, N., Chater, N., & Brown, G. D. (2006). Decision by sampling. Cognitive psychology, 53(1), 1-26.

Stewart, N., Reimers, S., & Harris, A. J. (2015). On the origin of utility, weighting, and discounting functions: How they get their shapes and how to change their shapes. *Management Science*, 61(3), 687-705.

Takahashi, H., Fujie, S., Camerer, C., Arakawa, R., Takano, H., Kodaka, F., et al. (2013). Norepinephrine in the brain is associated with aversion to financial loss. *Molecular Psychiatry*, *18*(1), 3–4.

Takeuchi, H., Kawada, R., Tsurumi, K., Yokoyama, N., Takemura, A., Murao, T., et al. (2015). Heterogeneity of Loss Aversion in Pathological Gambling. *Journal of Gambling Studies*, *32*(4), 1143–1154.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–518.

Trueblood, J. S., Heathcote, A., Evans, N., & Holmes, W. (2019). Urgency, Leakage, and the Relative Nature of Information Processing in Decision Making. *BioRxiv*, 706291.

Tsetsos, K., Chater, N., & Usher, M. (2012). Salience driven value integration explains decision biases and preference reversal. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(24), 9659–9664.

Tsetsos, K., Moran, R., Moreland, J., Chater, N., Usher, M., & Summerfield, C. (2016). Economic irrationality is optimal during noisy decision making, *113*(11), 3102–3107.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. Psychological review, 121(2), 179.

Turner, B. M., Forstmann, B. U., & Steyvers, M. (2019). Joint models of neural and behavioral data. Springer International Publishing.

Turner, B. M., Schley, D. R., Muller, C., & Tsetsos, K. (2018). Competing theories of multialternative, multiattribute preferential choice. Psychological review, 125(3), 329.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323.

Vermeer, A. B. L., Boksem, M. A. S., & Sanfey, A. G. (2014). Neural mechanisms underlying context-dependent shifts in risk preferences. *NeuroImage*, *103*(C), 355–363.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: an empirical validation. *Memory & Cognition*, *32*(7), 1206–1220.

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response-execution bias from decision bias: Arguments for an additional parameter in Ratcliff's diffusion model. *British Journal of Mathematical and Statistical Psychology*, *63*(3), 539–555.

Walasek, L., & Stewart, N. (2015). How to Make Loss Aversion Disappear and Reverse: Tests of the Decision by Sampling Origin of Loss Aversion. *Journal of Experimental Psychology: …*, *144*(1), 7–11.

Walasek, L., & Stewart, N. (2019). Context-dependent sensitivity to losses: Range and skew manipulations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(6), 957.

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 385–398.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*, *7*, 14.

Yechiam, E., & Hochman, G. (2013). Loss-aversion or loss-attention: The impact of losses on cognitive performance. *Cognitive Psychology*, *66*(2), 212–231.

Yechiam, E., & Telpaz, A. (2011). Losses Induce Consistency in Risk Taking Even Without Loss Aversion. *Journal of Behavioral Decision Making*, *26*(1), 31–40.

Zhao, W. J., Diederich, A., Trueblood, J. S., & Bhatia, S. (2019). Automatic biases in intertemporal choice. *Psychonomic Bulletin & Review*, 1–8.

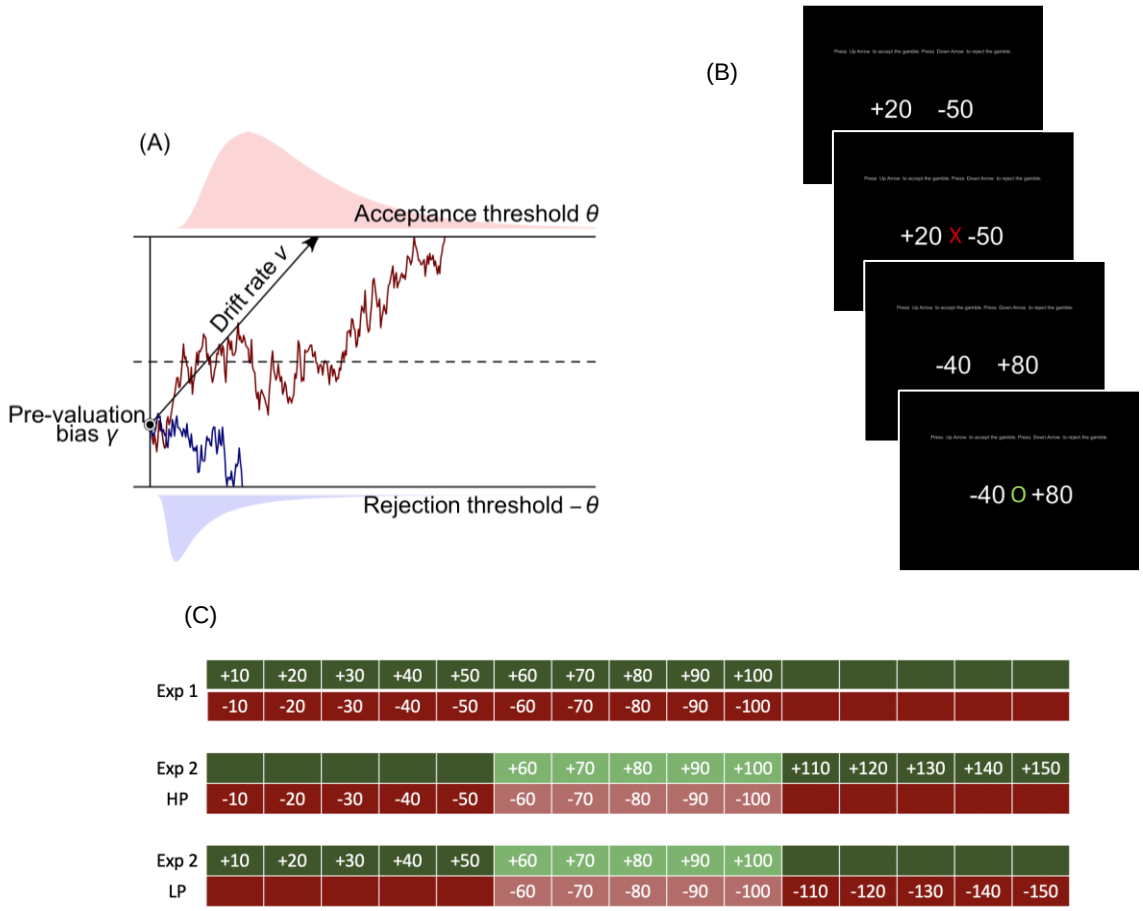


Figure 1. A: The drift diffusion model. The x-y axes represent time and accumulation state respectively. The arrow represents the expected accumulation process and its slope is the drift rate *v*, which corresponds to gamble utility. Each trajectory represents a hypothetical decision trial that starts from γ and ends when one of the thresholds, θ or $-\theta$, is hit. Note that γ ranges from -1 to 1, and $\gamma = 0$ indicates a neutral starting point. When $\gamma < 0$, decision makers have a pre-valuation bias to reject gambles, and the RTs for choices in which gambles are rejected are shorter than those in which gambles accepted (as indicated by the two RT distributions in the figure). B: Examples of experimental stimuli. For each gamble, token values of gains and losses were displayed side-by-side in counterbalanced positions. Participants used keyboards to indicate their responses, and the specific key-response association was held constant in a block (50 trials). Participants' responses were followed immediately by the appearance of a sign in the center of

the screen (red crosses for rejections and green circles for acceptances). (C) Gain (green) and loss (red) amounts used in the experiments. Lighter colors indicate shared gambles in both the high-payoff (HP) and low-payoff (LP) conditions in Experiment 2.

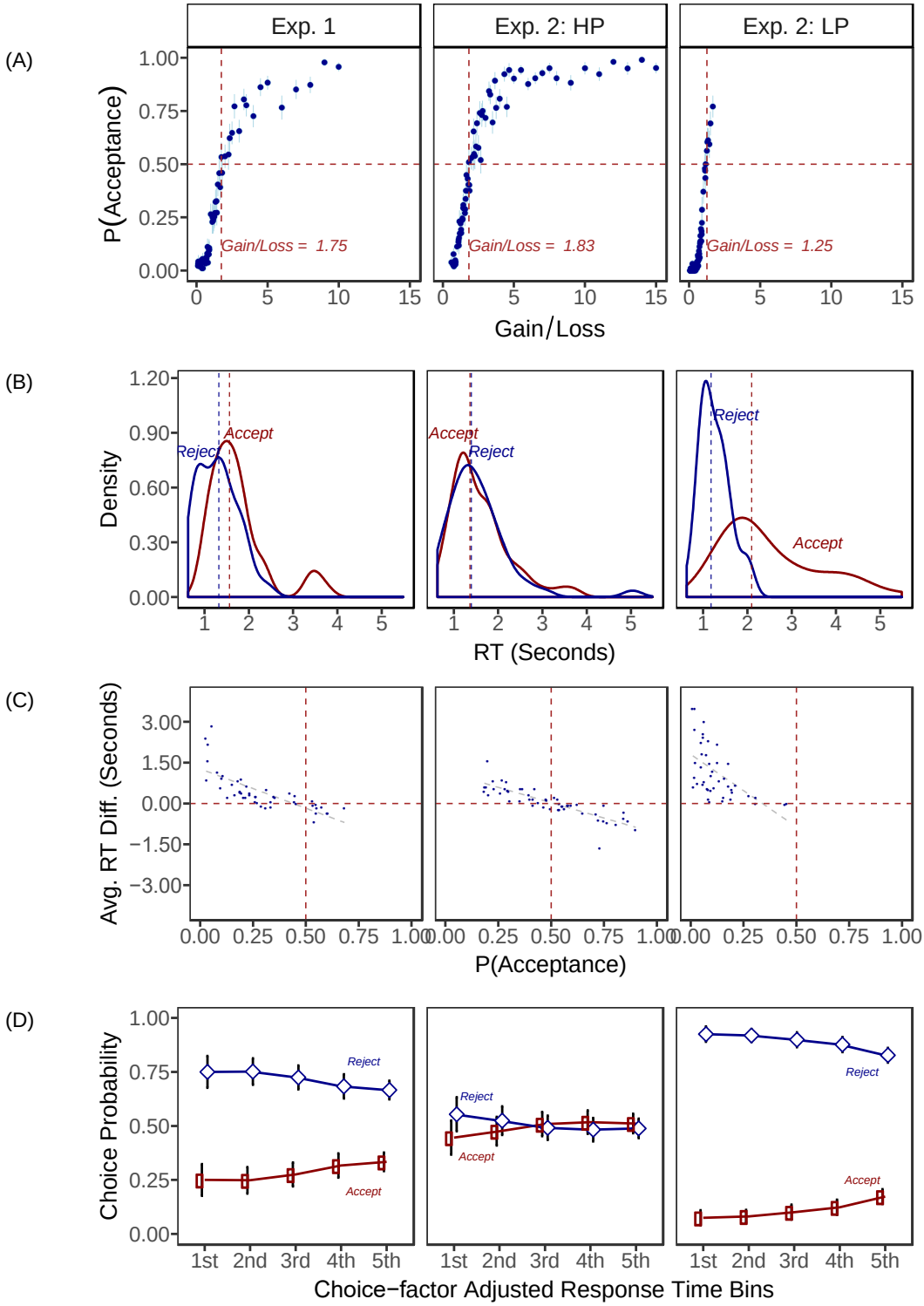
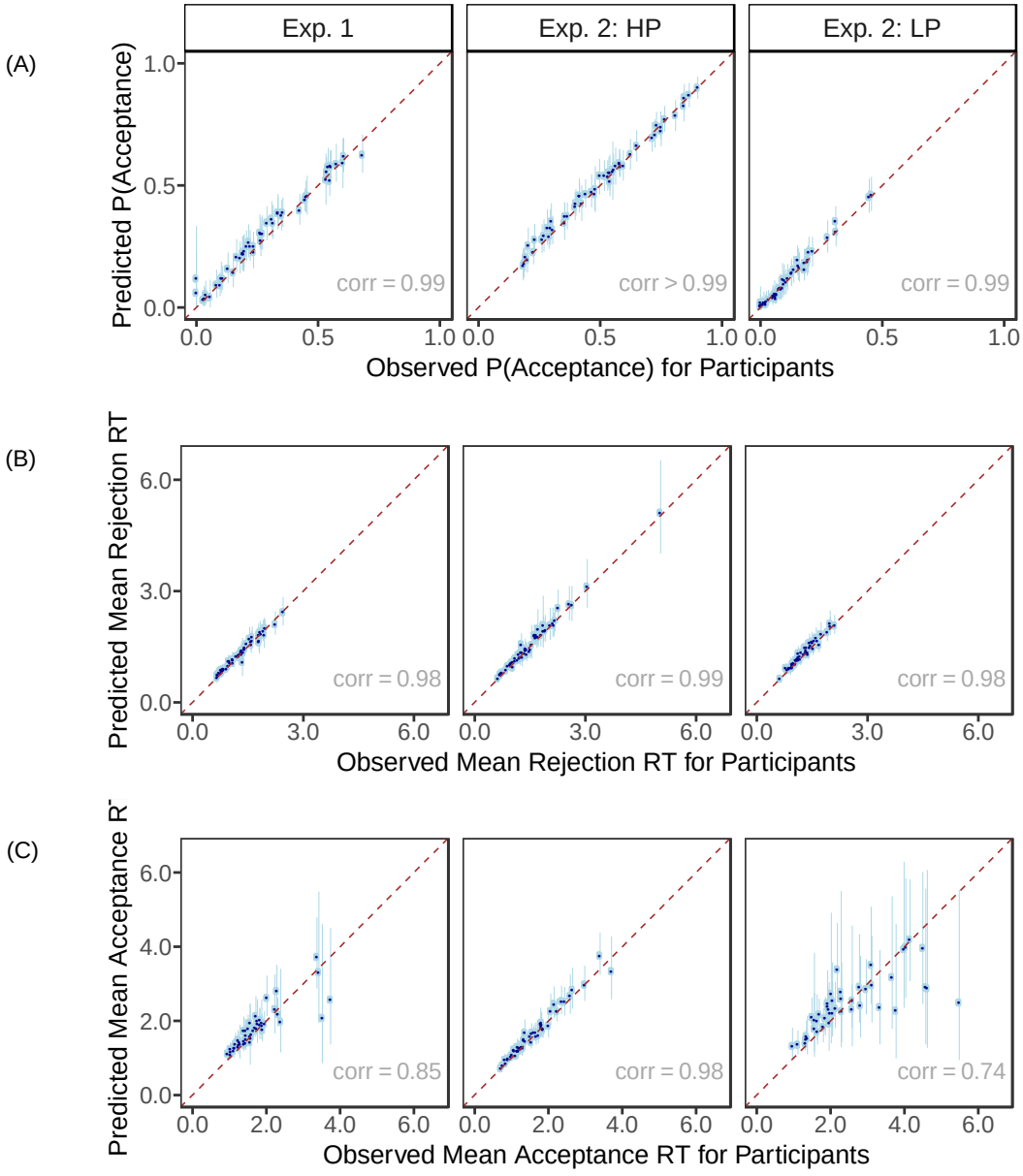


Figure 2. A: Average acceptance rates and the ratio of gains to losses, for each of the gambles used in the experiments. Error bars indicate standard errors across participants. B: Distributions of response times for acceptance and rejection choices in aggregate data. Dashed lines indicate medians of response times. C: Acceptance rates and RT differences for acceptance vs. rejection for each participant. D: Choice-RT relationships. Here the horizontal axis indicates RTs, which are adjusted for choice factors (gain and loss values) before being sorted into 5 bins. Trials with smaller (longer) adjusted RTs are on the left (right). The vertical axis indicates probabilities for rejection (blue diamonds) and acceptance (red circles). Error bars indicate 95% CI.



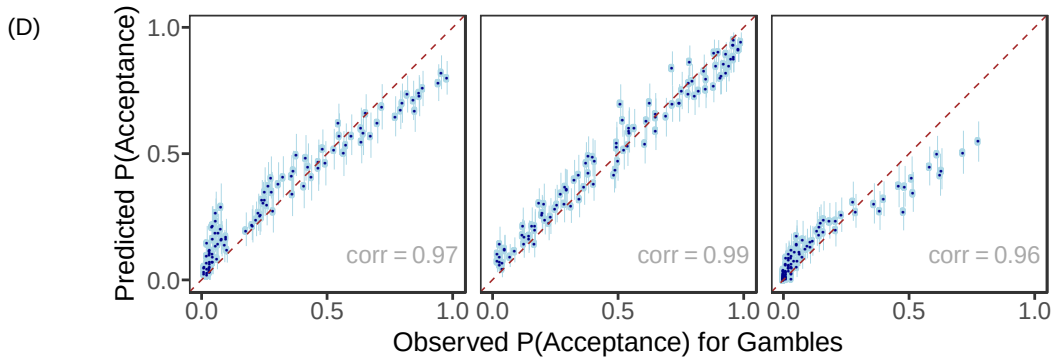


Figure 3. Posterior predictive check for model fits. A: Acceptance rates for participants. B: Mean RT for rejection decisions for participants. C: Mean RT for acceptance decisions for participants. D. Acceptance rates for gambles. In panels A-C, each dot represents a participant. In panel D, each dot represents a unique gamble. Error bars indicate 95% credible intervals. Corr corresponds to the Pearson correlation coefficient.

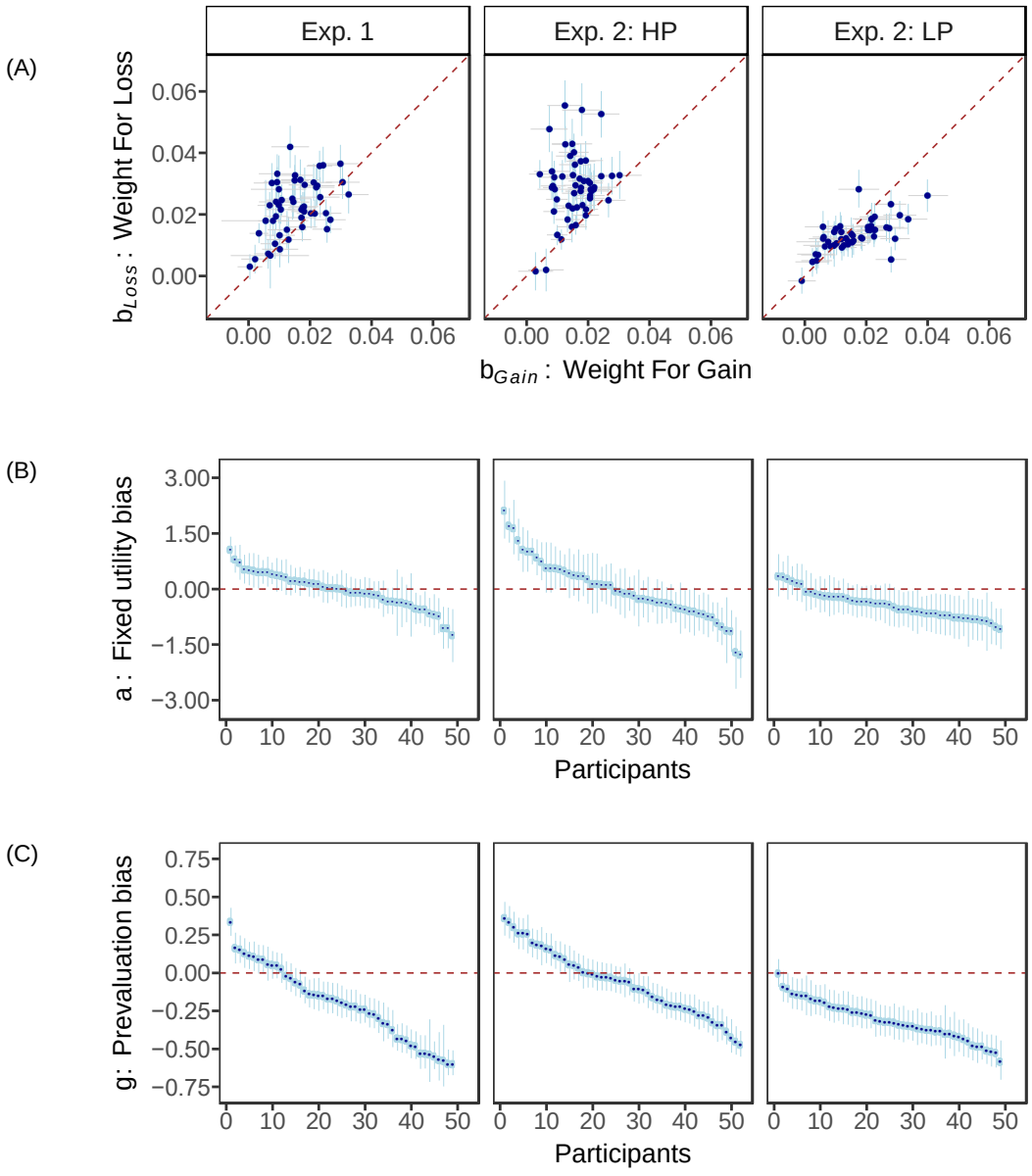


Figure 4. Participant-level estimates for the four DDM parameters. A: Prospect theory utility weighting bias. B: Fixed utility bias. C: Pre-valuation bias. Each dot represents a participant. Error bars indicate 95% credible intervals.

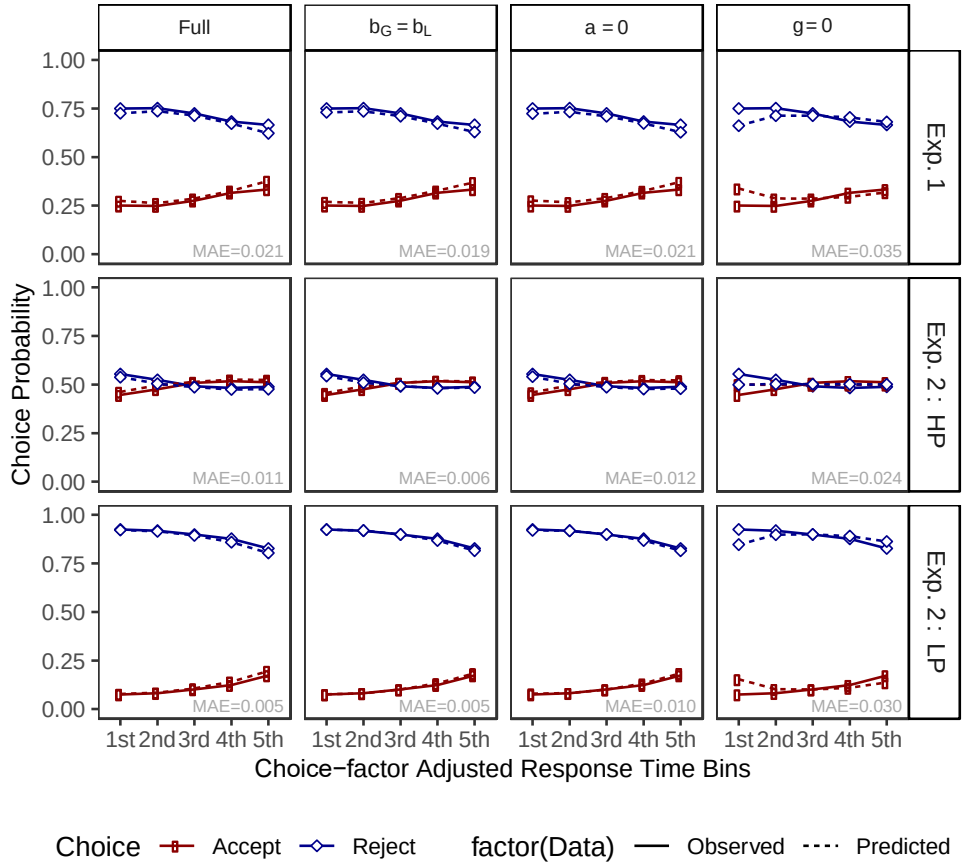


Figure 5. Choice-RT relationships for observed data (solid lines) and model simulated data (dashed lines) for the full model and for models restricting each of the three mechanisms. MAE: Mean absolute error.

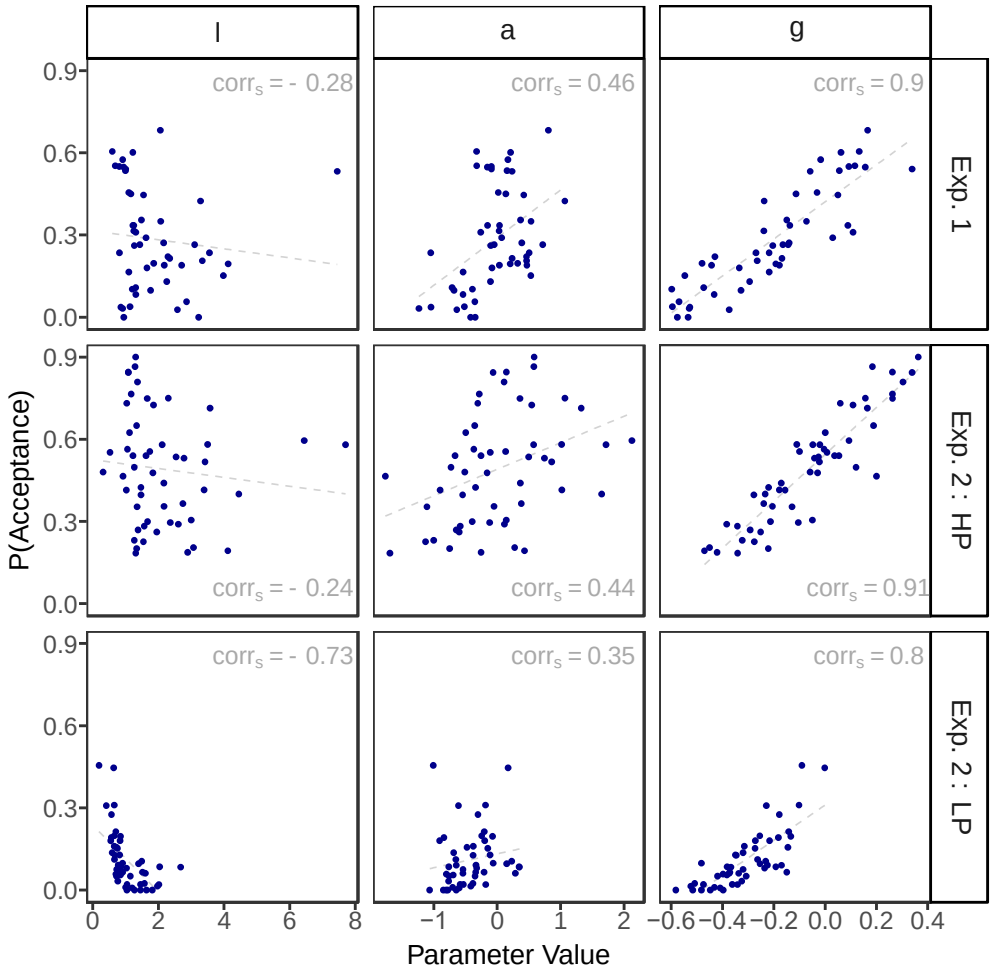


Figure 6. Relationships between the estimated DDM parameters and acceptance rates. Each dot represents a participant. $corr_s$: Spearman's correlation coefficients.

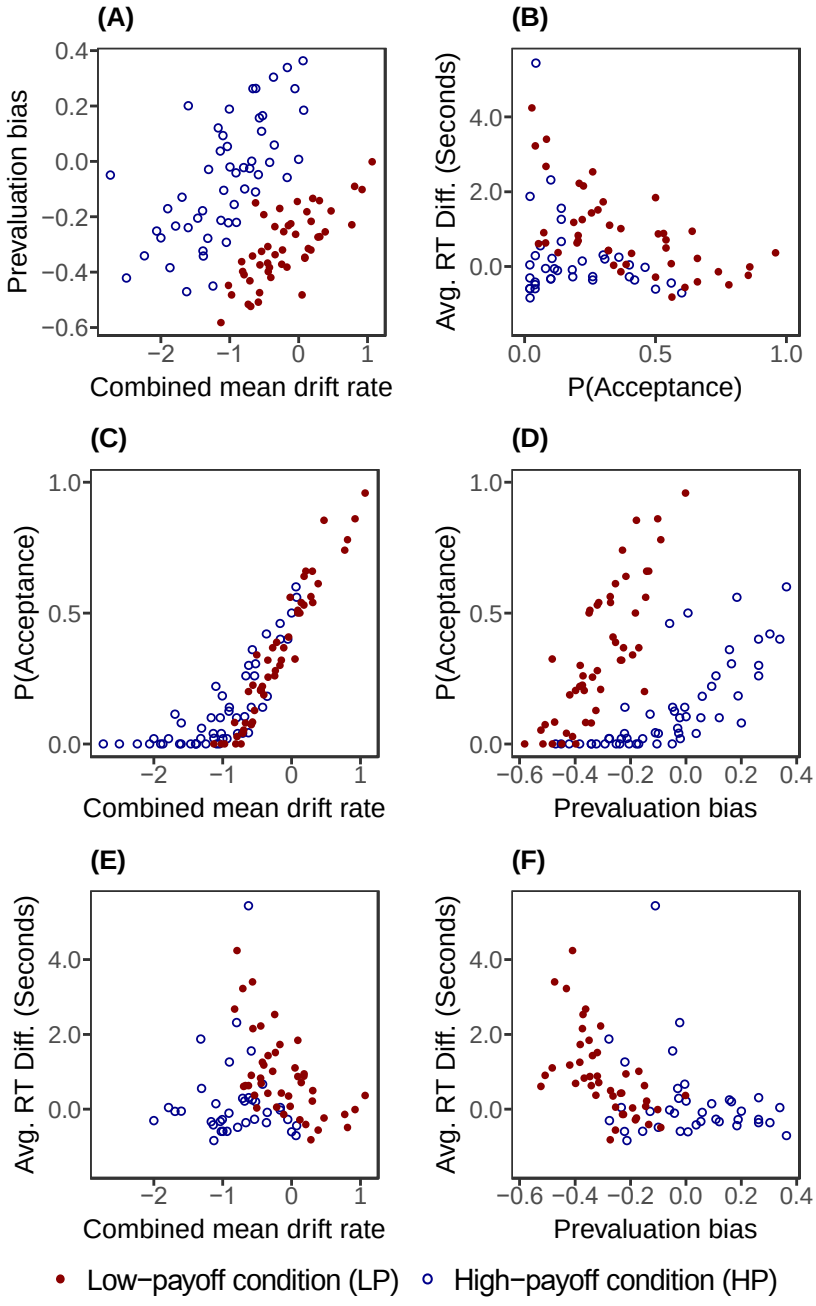


Figure 7. Analyses of shared gambles in Experiment 2. A: Disassociation between drift rates and pre-valuation biases across the HP and LP conditions. B: Disassociation between choice probabilities and RT differences across the HP and LP conditions. C: Effect of drift rates on acceptance probabilities. D: Effect of pre-valuation biases on acceptance probabilities. E: Effect of drift rates on averaged RT differences between acceptance and rejection decisions. F: Effect of pre-valuation biases on averaged RT differences between acceptance and rejection decisions. Solid red dots indicate LP condition and empty blue dots indicate HP condition.

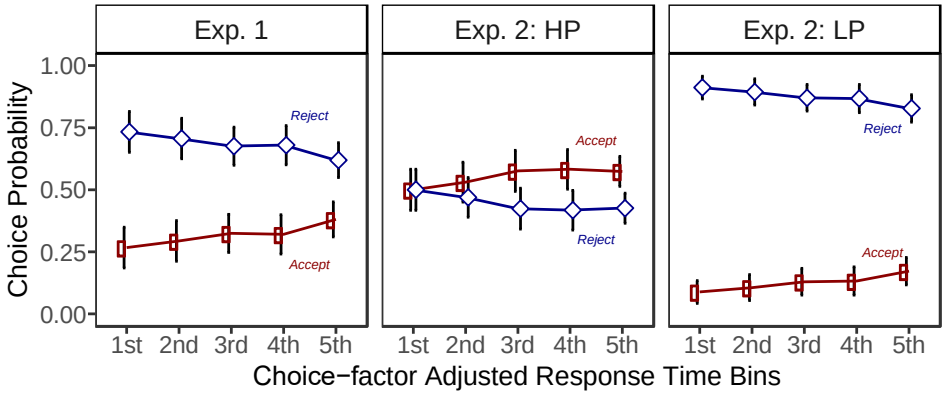


Figure 8. Choice-RT behavioral marker for the first 25 trials (first half of the first block). Participants already display a pre-valuation bias towards rejection at the very beginning of the experiment.

Table 1. Psychological mechanisms underlying loss aversion.

Mechanism (DDM parameter)	Interpretation	Features	Implication
Pre-valuation bias $\gamma < 0$	A predisposition towards rejection, corresponding to a biased prior expectation favoring rejection	Stimulus-independent Before the valuation process	Higher choice probabilities for rejection, and quicker RTs in rejection vs. acceptance decisions
Utility weighting bias $\lambda = \frac{\beta_L}{\beta_G} > 1$	Larger effect of losses relative to gains on utility (standard prospect theory explanation)	Stimulus-dependent During the valuation process	Higher choice probabilities for rejection, but no difference in RTs in rejection vs. acceptance.
Fixed utility bias $\alpha < 0$	Additive, bias in utility favoring rejection	Stimulus-independent During the valuation process	Higher choice probabilities for rejection, but no difference in RTs in rejection vs. acceptance

Table 2. Group-level parameters posterior distributions.

		$\boldsymbol{\beta_L}$	$\boldsymbol{\beta_G}$	$\boldsymbol{\alpha}$	$\boldsymbol{\gamma}$	$\boldsymbol{\theta}$	τ
Exp. 1	Mean	0.022	0.015	-0.032	-0.214	1.182	0.423
	Median	0.022	0.015	-0.031	-0.214	1.181	0.423
	2.5%	0.02	0.013	-0.207	-0.268	1.12	0.393
	97.5%	0.025	0.018	0.141	-0.161	1.248	0.456
	SD	0.001	0.001	0.089	0.027	0.033	0.016
Exp 2: HP	Mean	0.029	0.016	0.014	-0.065	1.226	0.422
	Median	0.029	0.016	0.016	-0.065	1.225	0.422
	2.5%	0.026	0.014	-0.246	-0.113	1.141	0.397
	97.5%	0.032	0.018	0.274	-0.017	1.317	0.448
	SD	0.002	0.001	0.133	0.025	0.045	0.013
Exp. 2: LP	Mean	0.013	0.016	-0.420	-0.316	1.412	0.397
	Median	0.013	0.016	-0.418	-0.316	1.411	0.396
	2.5%	0.011	0.013	-0.583	-0.354	1.333	0.370
	97.5%	0.015	0.019	-0.259	-0.278	1.496	0.425
	SD	0.001	0.001	0.082	0.019	0.041	0.014

Table 3. DICs of the full model and the three constrained models. The model eliminating the pre-valuation bias has large DIC increase from the full model.

	Full	Utility weighting constrained ($\boldsymbol{\beta_L = \beta_G}$)	Fixed utility constrained ($\boldsymbol{\alpha = 0}$)	Pre-valuation constrained ($\boldsymbol{\lambda = 0}$)
Exp. 1	17,184	17,548	17,372	18,141
Exp. 2: HP	20,886	21,957	21,249	21,655
Exp. 2: LP	16,233	16,419	16,362	17,147

Table 4. Standardized regressions of participant-level acceptance rates on the utility weighting bias, the fixed utility bias, and the pre-valuation bias. Note that the pre-valuation bias has the largest coefficients and thus correlates with participant acceptance rates the most across all experiments.

	Utility weighting (λ)	Fixed utility (α)	Pre-valuation (γ)	R^2
Exp. 1	-0.013	0.043**	0.148***	0.804
	[0.015]	[0.016]	[0.014]	-
Exp. 2: HP	-0.066***	0.098***	0.152***	0.908
	[0.016]	[0.016]	[0.011]	-
Exp. 2: LP	-0.039**	0.020†	0.059***	0.673
	[0.013]	[0.011]	[0.012]	-

Note. † indicates $p < .1$; * indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$. Values in brackets indicate standard errors for regression coefficients.

Table 5. Standardized regressions of participant-level acceptance rates on utility weighting and fixed utility biases in the constrained DDM eliminating the pre-valuation bias.

	Utility weighting (λ)	Fixed utility (α)	R^2
Exp. 1	-0.025	0.124***	0.468
	[0.020]	[0.020]	-
Exp. 2: HP	-0.158***	0.220***	0.736
	[0.019]	[0.019]	-
Exp. 2: LP	-0.056***	0.043**	0.336
	[0.013]	[0.013]	-

Note. † indicates $p < .1$; * indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$. Values in brackets indicate standard errors for regression coefficients.